# Impact of Content-Focused and Practice-Based Professional Development Models on Elementary Electric Circuits Teaching and Learning

Principal Investigators:

Joan I. Heller, Heller Research Associates

Judith Warren Little, University of California, Berkeley

Mayumi Shinohara, WestEd

Final Report

December 30, 2010

This material is based upon work supported by the National Science Foundation under Grant No. 0545445.

## Table of Contents

# List of Tables

## List of Figures

## Acknowledgments

The research team would like to acknowledge colleagues who made the study possible from the early design phases to the final analyses.

First, we thank the site coordinators who made the program implementation and teacher data collection possible: Kim Bess (San Diego), Marlene Kotelman (Tucson), Cindy Moss (Charlotte Mecklenburg), Carol Mueller (Huntsville), Darlene Ryan (Research Triangle Park), Andrew Schwebke (South Sound LASER), Mayumi Shinohara (San Francisco), and Meg Watson (Boston). We are also grateful to the numerous course facilitators who so skillfully delivered the professional development courses, working diligently to stay true to the specifications of each course in the sequence they were assigned. The participating teachers and their students around the country contributed the data on which this entire study depended. We recognize the burden associated with taking part in a research study of this magnitude and thank them for their time, commitment, diligence, and interest over the past several years.

This study was conducted jointly by researchers from Heller Research Associates, WestEd, and the University of California at Berkeley. Colleagues at our three organizations worked together for years as the research team developed the interventions, provided them to teachers, and collected data. We acknowledge the unwavering commitment of the implementation team and the entire staff that supported the project: Kirsten Daehler, Michelle Simone, Jennifer Mendenhall, and Mikiya Matsuda. We extend huge thanks to Cara Peterman at Heller Research Associates for her dedicated and diligent coordination of the data collection logistics from beginning to end. We thank Alissa Shethar, also of Heller Research Associates, for assisting in all aspects of instrument development and data collection, as well as Carol Verboncoeur, Alyson Spencer-Reed, and Rebecca Brown for their help with the instruments, data management, and project administration.

Throughout the study, we benefited from the expertise and technical assistance provided by Nicole Wong and Andrew Falk from UC Berkeley, who assisted in developing observation and interview protocols, and played an active role in collecting the classroom data.

## Executive Summary

This report summarizes a collaborative project by researchers at Heller Research Associates, WestEd, and the University of California at Berkeley. These researchers conducted a cluster-randomized experiment in eight sites across the United States over a two-year period to investigate three variants of a professional development model designed by WestEd's Understanding Science project. The sites involved 49 districts, more than 260 elementary teachers, and nearly 7,000 students, largely from underserved populations, including some classrooms in which 100 percent of students were eligible for free and reduced lunch and 65 percent were English learners.

The interventions were variants of an Understanding Science professional development course on electric circuits. The experiment contrasted three professional development models, all of which share the same core science investigation activities and science content focus, but otherwise vary the sources of science teaching knowledge (pre-structured written cases in Teaching Cases course; concurrent teaching and analysis of student work in Looking at Student Work course; metacognitive reflection on the teachers' own science learning in Content Immersion course).

Tests of electric circuits content knowledge were administered to all teachers at the beginning and end of the 2007–08 academic year and at the end of the following academic year, and to their students before and after the classroom electric circuits units during both 2007–08 and 2008–09. Because there were no off-the-shelf tests available, teacher and student tests were created by the research staff and validated for use in previous evaluations of the Understanding Science course on electric circuits. These tests were aligned with the Understanding Science project content framework, which specified the targets of instruction based on National Science Education Standards (National Research Council, 1996); Benchmarks for Science Literacy (American Association for the Advancement of Science, 1993); a host of state content standards; and frequently used, kit-based student science curricula, such as Full Option Science System (FOSS), Science and Technology Concepts (STC), and Curriculum Integration Program (CIP). The tests included questions reflecting the format and content of items in the Trends in International Mathematics and Science Study (U.S. Department of Education, 2004) and National Assessment of Educational Progress (NAEP). In addition to the multiple-choice test items, the content tests included open-ended questions to elicit explanations and applications of science content. Additional questions were included on the teacher content test to measure pedagogical content knowledge.

In addition to test results, data were collected regarding various characteristics of participating teachers. A teaching background survey provided information on professional experience and perspectives on science teaching. A randomly selected subsample of teachers participated in pre- and post-professional development interviews designed to measure *pedagogical content knowledge* (discussed in depth in the section titled "Important of Pedagogical Content Knowledge") and was observed and videotaped twice teaching lessons on electric circuits. Data were also collected during two rounds of professional development course implementation.

Using a randomized experimental design, this study established that the three closely related but systematically varied professional development models brought about highly significant gains in teachers' and students' scores on multiple-choice tests of science content knowledge, well beyond

those of comparable control groups, and that effects of the courses were maintained a year later. The score increases for students of intervention teachers occurred across a wide range of English language proficiency, from those speaking very little or no English to intermediate English learners, as well as native and non-native fluent-English-proficient students, and student gains did not differ significantly based on sex or race/ethnicity. These findings suggest that all three of the courses have design features that are effective at preparing teachers to support their students' science learning. The three interventions had in common identical collaborative science activities that engaged teachers in investigating and applying elementary grade electric circuits content, and it is not surprising that the three raised teachers' and students' scores on multiple-choice electric circuits tests. This result suggests that important elements in professional development can be configured in a number of ways and still have beneficial effects. The presence of certain characteristics is essential for high quality professional development, but it is possible for professional development to embody them in a variety of effective ways.

However, the three courses differed in key design elements, particularly with respect to sources of pedagogical content knowledge, and their efficacy for producing key outcomes varied accordingly. On a second measure of teacher and student content knowledge, the quality of explanations and applications in response to open-ended test items, the three courses significantly raised teacher explanation scores in Year 1, but only Teaching Cases had significant treatment effects in the follow-up year. For students, in Year 1 only the Looking at Student Work course significantly raised science explanation scores. In that year, Looking at Student Work teachers taught the unit on electric circuits concurrently with taking the professional development course. This meant that the Year 1 students in their electric circuits lessons were completing explanation tasks that were assigned as part of their teachers' Looking at Student Work course shortly before the students took the content posttest, giving them a considerable advantage over students of the other two intervention groups. Interestingly, in the follow-up year, the long-term effects of the Teaching Cases course matched the effects of the Looking at Student Work course, with both significantly raising students' science explanation scores. In contrast, other than teachers' Year 1 results, the Content Immersion course did not raise teachers' or students' scores on science explanations and applications. This pattern likely reflects the fact that Teaching Cases and Looking at Student Work both included analysis of student work and attention to the importance of using classroom tasks that are open-ended enough to elicit useful information about students' conceptual understandings, whereas the Content Immersion course did not include these components.

Teaching Cases and Looking at Student Work also significantly raised teacher pedagogical content knowledge scores, whereas the Content Immersion course did not. The measure of pedagogical content knowledge used in this study scored most highly student actions (e.g, "I would have them build the circuit"), incorporation of instructional representations and activities to help students make sense of phenomena (e.g., "We would trace the flow of current through the wires and into the bulb"), and explication of the specific science learning goals that would be targeted. Teaching Cases and Looking at Student Work both included explicit attention to the science understandings embodied in student work, and in analyzing classroom instruction in relation to helping students improve their understandings. Therefore, the significant impact on pedagogical content knowledge only resulting from Teaching Cases and Looking at Student Work courses is to be expected.

Teachers' ratings of the courses clearly express their preference for the Teaching Cases and Looking at Student Work courses over Content Immersion, in that the focus on content to the exclusion of classroom and student artifacts and reflection limited their opportunities to improve their own practices.

Finally, we also showed that the large impact of the courses on teacher content knowledge only partially accounts for student outcomes. Teachers get something else out of the courses, additional pedagogical knowledge that changes their teaching practices. Our measure of teacher pedagogical content knowledge was found to be a significant predictor both of student multiple-choice test scores and student science explanation scores. When both teacher content knowledge and pedagogical content knowledge were included in the HLM models, content knowledge was a significant predictor of student performance on the multiple-choice test items, but pedagogical content knowledge was not. However, when student science explanation scores were modeled instead of multiple-choice scores, both teacher content knowledge and pedagogical content knowledge were significant predictors of student performance.

In terms of research design, the power in this study lies in the combination of several design elements. The study compared carefully designed professional development variants, with components in common and that differ, each of which represents a strong option of interest to policy makers. The study used a set of measures driven by a conceptual logic model of the professional development model's target outcomes, and implemented a rigorous randomized experimental design that permits inferences about causal relationships. Finally, rich qualitative data were collected in addition to the test data to illuminate processes and relationships underlying the quantitative patterns.

## Background and Purpose

Questions regarding how to combine content and pedagogy in teacher courses to produce the greatest impact on teaching and learning represent the key challenge that motivated the research reported here. This study was designed to investigate the differential effects of three systematically varied teacher courses for elementary science teachers. The three teacher courses, developed by the Understanding Science program at WestEd, were intended to improve students' science achievement by strengthening their teachers' science content knowledge with science content components in common, while enhancing their pedagogical content knowledge in three different ways.

The 2007 landmark report, "Taking Science to School: Learning and Teaching Science in Grades K–8," produced by the National Research Council, concluded that "well-designed opportunities for teacher learning can produce desired changes in their classroom practices, can enhance their capacity for continued learning and professional growth, and can in turn contribute to improvements in student learning" (Duschl, Schweingruber, & Shouse, 2007, pp. 306–07). A number of theoretical models of the logic underlying such effective professional development have emerged in the literature that can guide development and evaluation of teacher education efforts (e.g., Cohen & Hill, 2000; Hanssen, 2006; Heller, Daehler, & Shinohara, 2003; Scher & Reilly, 2009; Weiss & Miller, 2006). A fundamental assumption of these models is that there is a clear cascade of effects from features of the professional development to immediate effects on teacher knowledge, intermediate effects on teacher practice, and finally more distal effects on students. More specifically, in this model professional development is assumed to achieve its effects on student achievement by improving both teachers' content and pedagogical content knowledge in ways that change their teaching practices.

Research has yielded a growing body of empirical evidence that teacher professional development can strengthen student achievement (e.g., Blank, de las Alas, & Smith, 2007; Cohen & Hill, 2001; Fennema et al., 1996; Franke et al., 2001; Saxe, Gearhart & Nasir, 2001). However, there is little evidence about the more specific, component linkages in the logic model—such as how professional development features influence what and how teachers learn, or the impact of different aspects of teacher change on student outcomes (Borko, 2004; Desimone, Porter, Garet, Yoon, & Birman, 2002; Fishman, Marx, Best, & Tal, 2003; Garet, Porter, Desimone, Birman, & Yoon, 2001; O'Reilly & Weiss, 2007; Scher & Reilly, 2009). In addition to this need for greater understanding of the mechanisms by which professional development can most effectively improve teaching and learning, there is a significant disjuncture between what is known already about quality professional development and what is available to districts, especially those facing poor student achievement and inadequate teacher preparation. Generally, school districts do not have access to coherent, effective professional development programs and they lack the site-based expertise and science-savvy staff developers to provide them (Little 2006; Duschl, Schweingruber, & Shouse, 2007). The majority of existing teacher education and professional development programs do not result in significant gains in teachers' content knowledge; nor do they bring about meaningful changes in teachers' instruction or support their "full implementation" of curricula (Hill, 2007; Scher & O'Reilly, 2009; Weiss, Banilower, McMahon, & Smith, 2001). The gap between the need for and availability of effective teacher professional development courses shaped this study's goal of determining how to design courses that *do* strengthen both teachers' content and pedagogical knowledge in ways that in turn improve their students' knowledge of that content.

*Rationale for focusing on electric circuits.* Elementary grade physical science was chosen as the domain for this study for several reasons. According to the 2000 National Survey of Science and Mathematics Education (Fulp, 2002), elementary school teachers are underqualified to teach science. Forty percent of elementary school teachers surveyed had taken four or fewer semesters of science coursework, and only four percent had undergraduate degrees in science or science education. Inadequate science preparation is particularly prevalent in schools with large populations of culturally and linguistically diverse students (Hart & Lee, 2003; Spillane, Diamond, Walker, Halverson, & Jita, 2001). It is not surprising, then, that on the 2000 National Assessment of Educational Progress in science, only 28 percent of grade 4 students were judged to be at or above the proficient level, with lower proficiencies for African American (7 percent), Latina/o (11 percent), and Native American (19 percent) students (Rodriguez, 2004).

In general, teachers' knowledge is most limited in the physical sciences (Fulp, 2002) as opposed to biological and ecological. Physical science ranked at or near the bottom of five science subjects in a measure of college science instruction among participating teachers. Despite comprising a large segment of national and state science standards (e.g., California Department of Education, 2003; American Association for the Advancement of Science, 1993; National Research Council, 1996), physical science, and electric circuits in particular, is often taught minimally at the elementary level, in part because teachers do not feel confident in this subject area (Heller & Kaskowitz, 2004). The current study addressed the need to better prepare teachers in this domain of science, to enhance students' potential to make sizeable gains in science achievement.

The broad aim of the research was to identify empirically-validated principles on which teacher course design could be based, as well as an understanding of how teachers' participation in professional development impacts their knowledge and practice and in turn improves students' learning. The study incorporates the kind of randomized experimental design that has been encouraged in educational research (Cook, 2001; Boruch, DeMoya, & Snyder, 2002; Jacob, Zhu, & Bloom, 2010; Slavin, 2002) along with depth and rigor in its systematic use of observational data.

### *Importance of Strengthening Teachers' Content and Pedagogical Content Knowledge*

The world of work requires the use of skills learned in science, such as deep critical thinking, inquiry, problem solving, and teamwork. Science education is essential in closing the skills gaps and responding to the labor needs and shortages in the 21st-century workforce (Partnership for 21st Century Skills, 2008), especially given that job growth is taking place in professional occupations such as health care and education and in technical fields such as computing (Terrell, 2007).

Many states have responded by setting high standards for students' science learning. Yet, for students to attain these standards, their teachers not only need a strong grasp of the subject matter; they also must know "how to organize, sequence, and present the content to cater to the diverse interests and abilities of the students" (Barnett & Hodson 2001, p. 432). Teachers are a dominant factor affecting student academic achievements (Aaronson, Barrow, & Sanders, 2003; Duschl, Schweingruber, & Shouse, 2007; Hill, Rowan, & Ball, 2005; Rivers & Sanders, 2002). At the elementary school level in particular, teacher effects are especially strong in relation to student academic gain (National Research Council, 2000; Wright, Horn, & Sanders, 1997). Nowhere is this clearer than in the specific impact of teachers' inadequate subject matter knowledge in science (Brewer & Goldhaber, 2000; Monk, 1994; Monk & King, 1994; National Research Council, 2000; Rowan, Chiang, & Miller, 1997). It has been well documented that elementary grade teachers have minimal backgrounds in science (Fulp, 2002; Broughman & Rollefson, 2000), and not surprisingly, students' academic performance in science has suffered, especially among underserved groups.

Strong knowledge of subject matter is essential but not sufficient for effective teaching. Teachers need subject-specific knowledge about how to teach in particular disciplines, or pedagogical content knowledge (Redish, 1996; Shulman, 1986; Bransford, Brown, & Cocking, 2000). Pedagogical content knowledge includes information about paths that students typically traverse in order to achieve understanding, and sets of potential strategies for helping students overcome the difficulties that they encounter. Although research has shown that the development of pedagogical content knowledge depends upon strong content knowledge (Clermont, Krajcik, & Borko, 1993; Lederman, Gess-Newsome, & Latz, 1994; Smith & Neale, 1989), strong content knowledge alone is not sufficient.

> Pedagogical content knowledge is not equivalent to knowledge of a content domain plus a generic set of teaching strategies; instead, teaching strategies differ across disciplines. Expert teachers know the kinds of difficulties that students are likely to face; they know how to tap into students' existing knowledge in order to make new information meaningful; and they know how to assess their students' progress. (Brown, Bransford, & Nasir, 2000)

Substantial progress has been made in describing pedagogical content knowledge and how it influences classroom teaching (Geddis, Onslow, Beynon, & Oesch, 1993; Lederman & Gess-Newsome, 1992; Parker & Heywood, 2000; Smith & Neale, 1989). Despite considerable evidence regarding the importance of pedagogical content knowledge, little is known about the conditions that foster its growth (Baxter & Lederman, 1999; Loughran, Mulhall, & Berry, 2003; Magnusson, Krajcik, & Borko, 1999; Van Driel, Verloop, & De Vos, 1998). Vital questions about how best to impart this essential pedagogical content knowledge remain unanswered.

> The challenge in providing [disciplinary content knowledge] is to equip the future teacher with *both* the content knowledge *and* an understanding of the thinking of children in the subject area. . . . Each is a critical component for effective teaching. . . . In light of this dual requirement, is content knowledge best obtained in disciplinary courses, . . . or in . . . courses that emphasize effective teaching of the content of the discipline? When content and teaching methods are taught separately, are teachers able to bridge the two? When they are done together, is adequate attention given to the disciplinary content?" (Bransford, Brown, & Cocking, 2000, p. 267)

The project addressed these questions through a randomized trial to compare the efficacy of three versions of a single in-service professional development course that vary only in how they incorporate pedagogical content knowledge. By comparing variants of one course, keeping course content consistent across interventions and controlling many variables that could influence the outcomes, it was possible to isolate the effects of the manipulated variables. The strategy was to start with an existing teacher course that included major elements of effective professional development practice as identified in the research literature (e.g., Birman, Desimone, Porter, & Garet, 2000), and then create two other variants. Each of the variants was designed to represent one of the main approaches to professional development that are currently being chosen by policy makers at the district and state level, and also to comprise as strong as possible an exemplar of its kind.

### *Rationale for the Interventions*

The interventions implemented in this study—variants of the Understanding Science professional development courses for elementary science teachers—embody characteristics described in research literature on effective programs. A good deal is known about the most successful features of professional development as described in the literature, including that it incorporate in-depth exploration of science content; teacher curricula embedded in standards-based student curricula; and a process that offers teachers opportunities for professional dialogue and critical reflection (Banilower, Boyd, Pasley, & Weiss, 2006; Cohen & Hill, 2001; Desimone, Porter, Garet, Yoon, & Birman, 2002; Knapp, McCaffrey, & Swanson, 2003; Little, 2006; National Staff Development Council, 2001; Wilson & Berne, 1999).

In the context of a high need for effective professional development programs that address teachers' content knowledge in science, a recent National Research Council report (Duschl, Schweingruber, & Shouse, 2007) calls for comprehensive professional development programs that are "conceived of, designed, and implemented as a coordinated system" to support students' attainment of high standards. The Understanding Science professional development courses offer just this kind of program to benefit teachers and students in K–8 science. The authors could find no other published courses that offered the combination of features provided in the Understanding Science courses, which are intended to help K–8 teachers learn science content in combination with analyzing student thinking and instruction. Most other professional development books currently available deal with just one or two of these topics (for example, science content or teaching)—leaving teachers the task of knitting together the information they most need to do their jobs well.

The Understanding Science approach focuses on the intersection of knowledge about content and teaching—that is, on developing teachers' pedagogical content knowledge. The model is based on the premise that, to develop this specialized knowledge, teachers must have opportunities to learn science content knowledge in combination with analysis of student thinking about that content and analysis of instructional strategies for helping students learn that content (Duschl, Schweingruber, & Shouse, 2007; Shinohara, Daehler, & Heller, 2004; Shymansky & Matthews, 1993; Van Driel, Verloop, & De Vos, 1998). Previous empirical studies provide consistent evidence that this model is effective for improving student science achievement (Heller, Daehler, & Shinohara, 2003; Heller & Kaskowitz, 2004).

*Course structure.* The course structure is designed to move teachers through learning about key science concepts, literacy supports, classroom practices, and students' science ideas, drawing from research in the fields of adult learning and cognitive psychology. With this path in mind, Understanding Science courses have three main components:

- *Science Investigations*—hands-on, guided inquiries designed to help adult learners explore core science concepts and classic misconceptions
- *Teaching Investigations*—discussions of narrative teaching cases drawn from actual classroom practice that provide a way for teachers to examine instructional strategies and student thinking, explore alternative solutions, and rethink their own teaching
- *Classroom Connections*—opportunities for teachers to read about, reflect on, and discuss key science and literacy concepts and consider how these concepts pertain to their own work with students

The materials for each course include a *Facilitator Guide* that provides detailed yet flexible procedures, in-depth background information (e.g., descriptions of the underlying science and common misconceptions of teachers), guiding questions and charts for each whole-group discussion, and other tips for leading a successful course. An accompanying *Teacher Book* presents all the materials teachers need to participate in a course, such as teaching cases of actual classroom practice, handouts, and session reviews summarizing the key concepts and outcomes, including illustrations of the common but incorrect ways students think about related concepts.

*Addressing the needs of English learners.* Understanding Science courses also have a focus on literacy, designed to help teachers and their students build important skills for reading and making sense of science texts. This unique component is one reason the Understanding Science courses have the potential to be particularly effective with English learners. Science achievement for English learners lags well behind that of native English speakers in the United States (Torres & Zeidler, 2002). Understanding Science professional development courses are designed to help teachers support students' reading, writing, and talking in science, as a means of making sense of the science and to help students develop academic language proficiencies. A portion of each course session focuses teachers' attention on identifying and evaluating literacy supports that guide learning. For example, the course is intended to help teachers learn that in order to lead successful discussions about science ideas, they need to make data public, visual, and manipulable, so students can discuss data sets, make comparisons, and draw conclusions. Teachers also practice and are expected to become fluent in using the representations most commonly used to organize and display data in different science disciplines. The Understanding Science professional development focuses teacher attention on the purpose and utility of different representations that they may use them more effectively.

Understanding Science courses also model and provide first-hand experiences for teachers in ways of learning science that research suggests are effective for all students, but especially for English learners. For example, English learners can benefit greatly from inquiry-based science instruction: hands-on activities based on natural phenomena depend less on mastery of English than do de-contextualized textbooks or direct instruction by teachers; and collaborative, small-group work provides opportunities for developing English proficiency in the context of authentic communication about science knowledge (Lee, 2002; Lee & Fradd, 2001).

*Professional development logic model.* The logic model motivating this approach (Figure 1) describes the cascade of influences connecting teachers' experiences in Understanding Science courses to student outcomes. The theory of action underlying the approach posits that when the professional development is situated in an environment of collaborative inquiry—one that is rich in talk about scientific meanings in conjunction with a focus on student thinking and critical analysis of practice—this leads to increases in teachers' science content and pedagogical content knowledge. These outcomes for teachers in turn result in changes in classroom practices, such as increased accuracy of science representations and explanations and a focus on conceptual understanding. Classroom changes ultimately produce improvements in student achievement, including reduced achievement gaps for low-performing students and English learners.

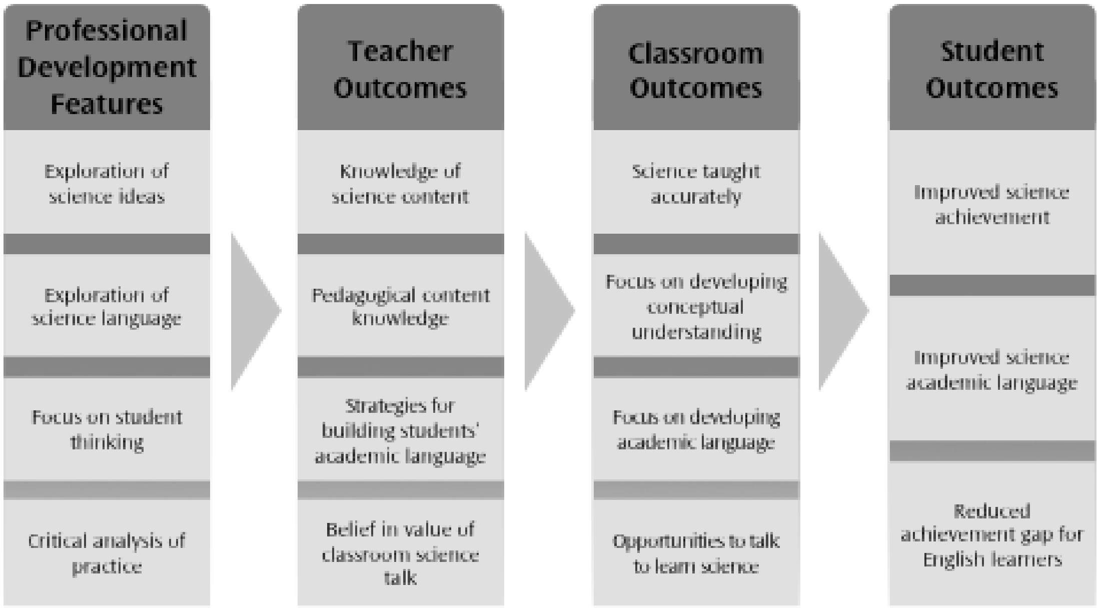


Figure 1. Understanding Science logic model
*Source* Developed by authors.

*Previous evidence of Understanding Science effects.* Over the past decade, evaluations of the Understanding Science professional development model has documented its effects on the science achievement of high-needs K–8 students, including English learners (Heller, Daehler, and Shinohara 2003; Heller & Kaskowitz, 2004). Quasi-experimental studies identified positive teacher and student outcomes of various Understanding Science courses for elementary and middle school teachers. While the non-experimental evidence did not allow definitive conclusions, the pattern of quantitative and qualitative findings suggests that gains were the result of teachers' participation in the Understanding Science courses. Findings of previous studies include:

- For teachers at both the elementary and middle school levels, in every study of various Understanding Science courses, differences between teachers' mean pre-course and post-course scores on science tests were statistically significant, with effect sizes from .44 to 1.09.
- At the elementary level, statistically significant differences favoring students in the

intervention group were found between the adjusted posttest mean for students of teachers who participated in Understanding Science courses (*n* = 123) versus the adjusted posttest mean for the comparison groups (*n* = 84) after controlling for pretest differences (*ES* = .84).

- English learners of intervention teachers (*n* = 97) made gains that were statistically significant, improving from pretest to posttest by 0.95 standard deviations more than comparison-group English learners (*n* = 57), with nearly a one standard deviation unit difference between the adjusted posttest mean for English learners in the intervention group and the adjusted posttest mean for English learners in the comparison group, favoring English learners in the intervention group.
- Students of all entering ability levels showed significant gains, with the greatest increase seen among low-performing students of intervention teachers (*ES* = 1.02).

In every study, statistically significant differences were found favoring intervention teachers and students on measures of science content knowledge (Heller, Daehler, and Shinohara 2003; Heller & Kaskowitz, 2004). Collectively, these data provide strong evidence of the internal validity of the professional development model. The data also offered promising indications of the model's impact on elementary-grade students' achievement across states, districts of varying sizes, with native English speakers and English learners, students with a range of socioeconomic backgrounds, and moderate evidence of such impact at the middle school level.

### ***Research Questions***

The research described in this report was designed to identify features of teacher education approaches that influence the impact of professional development on teachers' knowledge, classroom practice, and student achievement. At eight national sites with large populations of underserved students, the study investigated different ways to combine science content with pedagogy in teacher inservice courses on electric circuits at the elementary school level. In a randomized controlled experiment we compared three course designs, each modeled on an approach likely to be supported by districts or other teacher educators. Because the literature contains strong evidence of the critical role that content knowledge plays in the development of pedagogical content knowledge, all experimental course designs included an identical, strong content component in half of each course session. The three approaches varied with respect to the following sources of pedagogical content knowledge:

- primary emphasis on science content knowledge versus the inclusion of both science content and pedagogical content knowledge components, such as critical analysis of student work and teaching practices;
- the degree to which pedagogical issues related to teaching and learning the specific content are anticipated and controlled through pre-designed "cases" versus generated by the teachers; and
- the degree to which the experience of teaching students in the specific content domain, concurrently with professional development, becomes part of the professional development process.

### *Goal 1: Investigate the relative impacts of the three professional development interventions on target teacher and student outcomes.*

The experiment contrasts three models of professional development courses, all of which share the same core science investigation activities and science content focus, but otherwise vary the sources of science teaching knowledge (pre-structured written cases in Teaching Cases course; concurrent teaching and analysis of student work in Looking at Student Work course; content investigations and metacognitive reflection on the teachers' own science learning in Content Immersion course). All three approaches are based on the premise that, to develop pedagogical content knowledge, teachers must have opportunities to learn science content knowledge in combination with analysis of student thinking and instructional strategies related to that content (Shinohara, Daehler, & Heller, 2004; Shymansky & Matthews 1993; Van Driel, Verloop, & De Vos, 1998). The differences among the three allow the following questions to be addressed:

> ***Question 1:*** *What is the relative value of cases as resources for professional development, compared to the impact of teachers' analysis of their own concurrent classroom experience and student work?* [Teaching Cases vs. Looking at Student Work]
>
> ***Question 2:*** *What is the relative value of teachers' analysis of student work and related instructional issues during the professional development experience, over that of extended, in-depth immersion in science content?* [Teaching Cases vs. Content Immersion; Looking at Student Work vs. Content Immersion]
>
> ***Question 3:*** *What is the relative value of the three professional development interventions, compared to "business as usual" for elementary science teachers?* [Teaching Cases, Looking at Student Work, or Content Immersion vs. Control]

### *Goal 2: Examine the relationships among gains in teacher content knowledge, teacher pedagogical content knowledge, and student achievement.*

Classrooms are interactive places in which learning opportunities reside in the interplay of teachers, learners and content. This experimental study afforded an opportunity to examine relationships among teacher knowledge outcomes and student learning. Although research shows that teachers' content and pedagogical content knowledge are related to their teaching practices and student achievement, the dynamic interplay among teacher knowledge, practice, and student learning has not been studied, nor the role of teacher professional development in impacting these outcomes. The study quantitatively addresses the question:

> ***Question 4:*** Which, if either, is the better predictor of student achievement gains, teachers' post-course science content knowledge or pedagogical content knowledge?

We looked at these relationships quantitatively to determine whether teachers' post-course content or pedagogical content knowledge is a better predictor of student learning outcomes, and we are qualitatively documenting some of the reasons for the findings. Our speculation was that although teachers' content knowledge would impact classroom practice with respect to teachers' communication of sound, accurate, and grade-level-appropriate science content, changes in pedagogical content knowledge would be evidenced in more adaptive and effective instruction based on teachers' perceptions about students' understanding.

### *Goal 3: Examine the processes by which professional development designs achieve their effects on teacher knowledge and practice.*

Research on teachers' professional development has steadily converged on a small number of design features considered effective in enhancing teachers' subject knowledge for teaching (Garet et al., 2001; Cohen & Hill, 2001; Desimone, 2009). However, in most studies those features stand as proxies for professional development practice and for teachers' professional development experience. Few studies are designed to yield the kind of videotaped records that afford a close examination of professional development practice (among the exceptions, principally in the domain of mathematics, see Borko, 2006). Thus, we still know relatively little about the actual practices of professional development and how particular design features, as enacted, constitute opportunity for teacher learning. We have also only begun to investigate the role that facilitation expertise plays in scaling up professional development designs shown to be effective (Borko et al., 2010; Elliott et al., 2009). And we remain in the early stages of being able to link participation in professional development to changes in teaching and student learning.

This study capitalizes on a nested, multi-level research design to seek evidence of the relationships among professional development, change in teachers' thinking and practice, and student learning. In pursuit of this aim, this component of the analysis focuses on the implementation of the designed professional development (consistency of practice with design) and the ways in which selected activities, interactions, material resources, and strategic emphases may account for observed variation in teachers' measured science understanding and student learning gains. That is, this component of the research contributes to an explanation of the teacher and student outcomes by providing an analysis of systematic variations among different implementations of the professional development. For example, we examine how the specific practices of facilitation employed by expert and newly-trained facilitators during the science investigation open up or close off opportunities for participants to develop deeper conceptual understanding of the science by wrestling with new concepts, resolving confusion, reasoning about alternative formulations, and considering issues of classroom practice.

Analysis is organized by two main questions:

***Question 5***: How did the professional development courses, as implemented, reflect the intended design features? That is, how and to what extent did they engage teacher participation, achieve depth of talk about the key science ideas, elicit teachers' scientific reasoning, and enable focused about the teaching and learning of electric circuits content?

***Question 6***: What variations are evident in course implementation, and how do those variations relate to observed teacher and student knowledge outcomes?

## Professional Development Interventions

The study employed three experimental professional development course models, each of which was modeled after a different approach currently supported by districts and teacher educators. Each of the three course designs encompassed eight three-hour sessions completed over 8 to 14 weeks during the school year, or a five-day period during the summer. The courses focused on the teaching of electric circuits, a common component of elementary school curricula. All three course designs contained core features of effective science professional development, based on evidence in a growing body of literature (Banilower, Boyd, Pasley, & Weiss, 2006; Borko, 2004; Cohen & Hill, 2001; Garet et al., 2001).

Each course consisted of two portions—a science content component and a set of strategies intended to develop pedagogical content knowledge. Every course model involved teachers in a common set of science investigations that enabled in-depth, collective exploration of science content. However, the models varied from one another with regard to additional activities designed to support the development of pedagogical content knowledge. In Course A, *Teaching Cases*, teachers read and discussed written narratives containing student work and dialogue, teacher thinking, and descriptions of instructional materials and activities. Course B, *Looking at Student Work*, involved teachers in analyzing the affordances of specific classroom assessment tasks, and discussing evidence of students' scientific thinking drawn from their own classrooms. Course C, *Content Immersion*, combined the science investigations with a meta-cognitive learning analysis of the teachers' own science learning processes. Teachers in the Teaching Cases and Content Immersion courses taught electric circuits units following completion of the course; teachers in the Looking at Student Work course taught the unit concurrently with the professional development over a two-month period so as to supply samples of their own students' work.

The professional development interventions that were provided are variants of a course developed at WestEd with funding from the Stuart Foundation. The full set of *Understanding Science* courses will make up a comprehensive curriculum, including 15 courses on the major ideas of K-8 earth, life, and physical science. The course sessions focus on concepts in science in the context of narrative cases of practice drawn from actual classroom episodes involving those concepts (e.g., Barnett-Clarke & Ramirez, in press; Daehler & Shinohara, 2001; Merseth, 1996; J. Shulman, 1992; Wasserman, 1993). Written by classroom teachers, these case materials contain student work, student-teacher dialogue, context information, and teacher thinking and behaviors. Course sessions include hands-on science investigations that parallel those done by students described in the cases, building on research that shows teachers' knowledge growth results from professional development in which teachers encounter subject content through school curricula (Cohen & Hill, 2001; Saxe, Gearhart & Nasir, 2001).

Furthermore, *Understanding Science* courses were designed with features that are particularly well adapted for urban populations that are severely underserved with respect to science instruction (Hewson, Kahle, Scantlebury, & Davis, 2001; Spillane, Diamond, Walker, Halverson, & Jita, 2001). Research suggests that these students can benefit greatly from inquiry-based science instruction: (a) hands-on activities based on natural phenomena are more accessible to students with limited science experience, and depend less on mastery of English than do decontextualized textbook knowledge or direct instruction by the teacher; and (b) collaborative, small-group work provides opportunities for developing English proficiency in the context of authentic communication about science knowledge (Lee, 2002; Lee & Fradd, 2001; Rosebery, Warren, & Conant, 1992). The *Learning Science for Teaching* course in electric circuits was developed by urban teachers and field tested with ethnically, culturally, socioeconomically, and linguistically varied groups of students and teachers.

For the purposes of this study, a course in elementary grade science already existed that had undergone extensive refinement, from preliminary trials through pilot testing and national field tests. There were strong indications from three years of evaluation studies that the Understanding Electric Circuits course brings about desired teacher and student outcomes (Heller, Daehler, & Shinohara, 2003; Heller & Kaskowitz, 2004). In a pre-post quasi-experimental study, project teachers showed significant gains of more than one standard deviation on tests of content knowledge about electricity and magnetism (Heller & Kaskowitz, 2004), and important changes in pedagogical content knowledge as demonstrated through in-depth assessment interviews requiring reasoning about student work and instruction (Heller, Daehler, Shinohara, & Kaskowitz, 2004).

With respect to student achievement, a comparison group study (Heller, Daehler, & Shinohara, 2003) showed that scores on a test of electricity and magnetism increased significantly from pre- to posttest among students of teachers in the project. There were no significant pre-post gains among students of non-participating teachers who had also taught a unit on electricity and magnetism during the same period. In addition, adjusted posttest scores for the project group were significantly greater than for the comparison group after controlling for pretest differences. An especially encouraging result is that low-performing students of project teachers made the biggest gains, improving by more than two standard deviations, closing a portion of the achievement gap between students.

### *Course Materials*

To assist facilitators and ensure consistency of the intervention across sites, materials for the course were made available as a set of two books—a Participant Book for teachers and a Facilitator Guide for staff developers. The Participant Book contains five chapters (one per session) and presents all the materials teachers need to participate in the professional development course. Each chapter contains a teaching case that illustrates students' science thinking and highlights an important teaching dilemma that any teacher might face, along with a companion content guide that explains and illustrates core science concepts. In addition, each chapter includes science investigation handouts and case discussion handouts that guide teachers' small-group working time and structure their conversations around science, student thinking, and instruction.

The Facilitator Guide contains five core chapters (one per session) and provides extensive support materials and detailed procedures needed to successfully lead a course. Each chapter describes the underlying science (including common yet incorrect ideas children and adults have) and provides scripted yet flexible procedures, such as instructions to guide the hands-on and sense-making work in each science investigation, guiding questions for each case discussion, and instructions for helping teachers complete classroom connection assignments between sessions.

### *Course Features*

Each of the three interventions consisted of a 24-hour electric circuits course composed of eight three-hour sessions. Course sessions during the summer took place over one five-day week, and those during the school year were held every other week for 14 weeks. The three interventions vary the following sources of pedagogical content knowledge:

- *content* knowledge acquired through science investigations, written content guides, lectures from experts, and opportunities for sense-making discussions among the teachers;

- analysis of content-specific instruction, presented in *cases* crafted to capture classroom dilemmas growing from relationships among content, student thinking, and instructional strategies; or from teachers' first-hand experiences and samples of student work;

- insights from *teaching* the student curriculum (either previously or simultaneously) and bringing classroom observations and evidence to the group for reflection and analysis.

Because the literature contains strong evidence of the critical role that content knowledge plays in the development of pedagogical content knowledge, all experimental course designs included a strong content component. The content component for each course includes elements specific to pedagogical content knowledge. These include hands-on science investigations drawn from the classroom curriculum that teachers use with their students, facilitated sense-making discussions, and written content guides. By combining this content component in different ways with case discussions and concurrent teaching of the curriculum, we generated the four variants listed in Table 1.

Table 1. Sources of Pedagogical Content Knowledge in Four Experimental Conditions

| | Source of pedagogical content knowledge | | | |
|---|---|---|---|---|
| Experimental condition | Science content | Written teaching cases containing student work | Concurrent teaching and teacher-selected student work[a] | Metacognitive analysis of own learning |
| Teaching Cases | X | X | | |
| Looking at Student Work | X | | X | |
| Content Immersion | X | | | X |
| Control | | | | |

[a] Teachers in the Teaching Cases, Content Immersion, and Control groups brought to the courses whatever previous experience they had had teaching electric circuits, and taught the unit within two months after the end of the course.
*Source* Developed by authors.

Each experimental condition included a science investigation component that is common to each intervention. In addition, participants in the first model, *Teaching Cases*, read and discussed pre-defined cases while participating in the professional development. Participants in the second model, *Looking at Student Work (LASW)*, taught the unit concurrently with the professional development and engaged in structured analysis of their own teaching and student work. Those in the third model, *Content Immersion*, engaged in reflective discussions about their own learning processes that occurred during the science investigation segment of each session. Finally, a *Control* condition took no project professional development until after data collection in Rounds 1 and 2 was completed. The control teachers were free to participate in any inservice courses they would ordinarily take during the school year.

In terms of the ways that the models were instantiated in actual courses, the experimental variations in Table 2 were implemented according to the time allotments shown in Table 3. Each session of each course variant had some of the following components:

- science investigation and structured sense-making discussion of the content (in all variants);
- written case and case discussion (Teaching Cases);
- structured discussion analyzing participants' own teaching and student work (Looking at Student Work)
- structured reflection on the teachers' own processes of learning during the science investigation (Content Immersion).

Table 2. Opportunities to Develop Pedagogical Content Knowledge in Three Professional Development Models

| Aspects of pedagogical content knowledge | Features of professional development |
|---|---|
| Science content | Guided science inquiry, written content guides, and opportunities for sense-making discussions among the teachers (All) |
| Content-specific instructional practices | Instruction presented in teaching cases crafted to capture classroom dilemmas growing from relationships among content, student thinking, and instructional strategies (Teaching Cases)<br>or<br>Instruction drawn directly from teachers' first-hand experiences (Looking at Student Work) |
| Insights into student thinking | Student work presented in teaching cases designed to reveal major ways that students think about the content, both correct and incorrect (Teaching Cases)<br>or<br>Student work drawn from teachers' current classes (Looking at Student Work)<br>or<br>Metacognitive analysis of own thinking and learning (Content Immersion) |

Table 3. Time Allotted in Each Course Session to Components of Professional Development Models

| | Structure of each course session | | | |
|---|---|---|---|---|
| Experimental condition | Science investigation | Analysis of student thinking and instruction in teaching cases | Analysis of own students' thinking and instruction (concurrent teaching) | Metacognitive analysis of own thinking and learning |
| Teaching Cases | 1 hour | 2 hours | 0 | 0 |
| Looking at Student Work | 1 hour | 0 | 2 hours | 0 |
| Content Immersion | 1 hour | 0 | 0 | 2 hours |

*Source* Developed by authors.

Teaching Cases is the most fully integrated model in that it combines content and pedagogy in multiple ways, throughout each course session. Content Immersion provides the greatest contrast to Teaching Cases in relation to research question 1 regarding how best to combine content and pedagogy, while Teaching Cases and Looking at Student Work represent our attempt to get at the independent contribution of pre-designed cases versus analysis of student work generated during one's own teaching of the curriculum. Content Immersion is included because it most closely corresponds to a science immersion approach, which is often the choice of administrators and policy makers. It should be noted, however, that this variant goes beyond a strictly content-only approach, such as that in a college science course, by including science investigations that situate teacher content learning in student curriculum materials and activities, and in its metacognitive component.

### *Facilitator Selection and Training*

Site coordinators and district staff at each site helped identify and solicit the participation of professional development leaders who might facilitate the courses. Understanding Science staff selected the facilitators from among those individuals through telephone interviews with candidates. During the selection process, staff considered the following qualifications:

- At least two years' experience leading teacher professional development courses in elementary science.
- Strong content knowledge in physical science, ideally in the specific content topic of the professional development course.
- Experience teaching that content to the grade addressed in the study.
- Strong pedagogical content knowledge, including ability to describe what tends to be difficult for students and teachers to understand about electric circuits, and ability to generate instructional strategies that address specific learner misconceptions.
- Good fit with the Understanding Science professional development model, including a social constructivist perspective focusing on helping students and teachers learn through collaborative discourse about science.
- Ambitious expectations for learners.

- Acceptance of and commitment to following a strict professional development and research protocol for the larger good in science education.
- Promptness in responding to email.

Two facilitators were selected from each site with the exception of the San Francisco Bay Area, where two WestEd Understanding Science program staff served as facilitators. The facilitator pairs at each site were assigned to one of the sequences of two courses at that site (see Table 3) with the exception, again, of the San Francisco Bay Area, where one program staff person facilitated each course.

Facilitators received training before each of the two implementation rounds. In a five-day facilitator orientation and training held prior to Round 1 (July 2007), the 10 newly-recruited facilitator learned in general terms about the research goals and the three experimental professional development models, and were told which sequence of courses they were assigned to teach. Those who were about to facilitate the Looking at Student Work and Content Immersion groups learned, for example, that some of the courses include a case discussion component, but in the training they did not read or work with the cases. They were then trained separately in the course components of the model to which they had been assigned for the first round. After the first round, a second training was held (December 2007) to prepare facilitators for the next course variants they would be teaching. Researchers observed and videotaped the trainings to document how the three models were presented.

The training distinguished between a focus on science content, and attention to pedagogical concerns related to that content. The research was presented as an investigation about the tradeoffs among various ways of combining the two. Facilitators then experienced the professional development intervention themselves, completing two electric circuits course sessions per intervention over three days. The majority of the training time was spent deepening facilitators' understanding of electric circuits, grounding them in the common yet incorrect ideas students (and adults) have about the science, and helping participants develop the necessary facilitation skills. Project staff modeled facilitation, engaged the group in analyzing video clips of exemplary facilitation, and provided the trainees with practice in facilitating at least one course session. Facilitators used the course materials—the Facilitator Guide and Participant Book—throughout the training.

The group training segments on each course also allowed project staff to confront potential facilitator bias. In this case, as facilitators experienced the different interventions, they could have believed that one is better than the others. Conversations to bring such biases to their awareness were included in the training as needed, to help facilitators understand the logic underlying each model and stay true to the model they were leading at any given time.

### *Course Implementations*

In summer 2007, teachers who were randomly assigned to the intervention group took the Understanding Electric Circuits course, led by pairs of trained facilitators at each site. The two facilitators for each course alternated between serving as lead facilitator and serving as co-facilitator for each session. As shown in Table 4, an average of just over 60 percent of teachers initially assigned to the intervention group received the intervention, ranging from 40 percent to

75 percent at individual sites. At the time of the intervention, however, some teachers were no longer eligible to take the course, either because their school or district did not agree to participate in the study or because they had left teaching or moved to a different grade or school.

Table 4. Number of Intervention Group Teachers Initially Assigned and Actually Participating in Summer 2007 Understanding Science Courses, by Research Site

| Site | Number of initially assigned intervention teachers | Actual number of teachers completing courses | Percent of initially assigned teachers that completed courses |
|---|---|---|---|
| 1 | 30 | 12 | 40.0 |
| 2 | 27 | 17 | 63.0 |
| 3 | 15 | 10 | 66.7 |
| 4 | 45 | 34 | 75.6 |
| 5 | 59 | 36 | 61.0 |
| 6 | 54 | 30 | 55.6 |
| 7 | 56 | 34 | 60.7 |
| 8 | 38 | 28 | 73.7 |
| Total | 324 | 201 | 62.0 |

*Source* Authors’ analysis of primary data collected for the study.

Attendance records were kept for each session of each course implementation. Overall, attendance rates were strong (see Table 5), with almost 95 percent of the teachers attending all, or all but one, three-hour session of the eight course sessions. The frequencies of missing more than two sessions varied, however, from 3–4 percent for Teaching Cases and Looking at Student Work to over 11 percent for the Content Immersion course. Reasons for the lower attendance rates will be considered later in the report.

Table 5. Number and Percent of Intervention Group Teachers Who Missed One or More Summer 2007 Understanding Science Course Sessions, by Experimental Condition

| | Teaching Cases | | Looking at Student Work | | Content Immersion | | Total | |
|---|---|---|---|---|---|---|---|---|
| Number of class sessions missed | Number | Percent | Number | Percent | Number | Percent | Number | Percent |
| None | 59 | 80.8 | 45 | 66.2 | 42 | 66.7 | 146 | 71.6 |
| 1 | 10 | 13.7 | 21 | 30.9 | 14 | 22.2 | 45 | 22.1 |
| 2 | 1 | 1.4 | 0 | 0.0 | 0 | 0.0 | 1 | 0.5 |
| 3 | 1 | 1.4 | 0 | 0.0 | 1 | 1.6 | 2 | 1.0 |

| Number of class sessions missed | Teaching Cases |  | Looking at Student Work |  | Content Immersion |  | Total |  |
|---|---|---|---|---|---|---|---|---|
|  | Number | Percent | Number | Percent | Number | Percent | Number | Percent |
| 4 | 0 | 0.0 | 0 | 0.0 | 2 | 3.2 | 2 | 1.0 |
| 5 | 0 | 0.0 | 2 | 2.9 | 2 | 3.2 | 4 | 2.0 |
| 6 | 1 | 1.4 | 0 | 0.0 | 2 | 3.2 | 3 | 1.5 |
| 7 | 1 | 1.4 | 0 | 0.0 | 0 | 0.0 | 1 | 0.5 |
| More than 2 | 3 | 4.1 | 2 | 2.9 | 7 | 11.1 | 12 | 5.9 |
| Total | 73 | 100.0 | 68 | 100.0 | 63 | 100.0 | 204 | 100.0 |

*Source* Authors' analysis of primary data collected for the study.

## Research Design and Methods

### *Design Overview*

The goal of this study was to compare the efficacy of three Understanding Science professional development courses, using a cluster randomized trial design with repeated measures, with three intervention groups and one control group. Teachers served as the unit of randomization, and students were nested within teachers. Teachers were randomly assigned to an intervention or control condition and remained in their assigned condition until the conclusion of the study. The trial was conducted over a two-year period at eight national research sites. The specific outcomes measured were closely aligned with features of the three interventions, which incorporated in different ways science content, analysis of student thinking, and discussion of issues related to teaching that content. The study relied on a combination of quantitative and qualitative measures to investigate the impact of each intervention on teachers' and students' knowledge of that content, on teachers' classroom practices, and on teachers' pedagogical content knowledge about teaching and learning of that content.

The experimental design in Table 6 shows the sequence of data collection events (O) and interventions (X) for teachers, students, and classrooms in two rounds and a follow-up, for the 2007-08 and 2008-09 school years. At each site during the first year, one or two of the three electric circuits courses were implemented during each of two rounds involving different cohorts of teachers. During the follow-up data collection in the next school year, participants included teachers in the three professional development courses and their cohort of students that year. Control teachers did not participate in the follow-up so they could receive the professional development intervention by the beginning of Year 2.

Table 6. Longitudinal Experimental Design With Four Experimental Conditions and Pre–Post–Follow-up Teacher and Student Measures

| | Random-ization | Round 1 Fall 2007 | Winter 2008 | Spring 2008 | Summer 2008 | Follow-up Fall 2008 |
|---|---|---|---|---|---|---|
| Teachers | | Round 1 teachers | | | | Round 1 teachers |
| Teaching Cases | R | $O_{ST}$ X $O_{ST}$ | $O_S$ | | | $O_{ST}$ |
| Looking at Student Work | R | $O_{ST}$ X $O_{ST}$ | $O_S$ | | | $O_{ST}$ |
| Content | R | $O_{ST}$ X $O_{ST}$ | $O_S$ | | | $O_{ST}$ |
| Control | R | $O_{ST}$ — $O_{ST}$ | $O_S$ | | X | - |
| Students | | | Round 1 student cohort | | | Follow-up student cohort |
| Teaching Cases | NR | - | $O_T$ $X_S$ $O_T$ | | | $O_T$ $X_S$ $O_T$ |
| Looking at Student Work | NR | $O_T$ $X_S$ $O_T$ | | | | $O_T$ $X_S$ $O_T$ |
| Content | NR | - | $O_T$ $X_S$ $O_T$ | | | $O_T$ $X_S$ $O_T$ |
| Control | NR | - | $O_T$ $X_S$ $O_T$ | | | - |
| Intensive substudy teachers/classrooms | | | | | | |
| Teaching Cases | | $O_I$ | $O_{ObVR}$ $O_I$ | | | |
| Looking at Student Work | | $O_I$ | $O_I$ | | | $O_{ObVR}$ $O_I$ |
| Content | | $O_I$ | $O_{ObVR}$ $O_I$ | | | |
| Control | | $O_I$ | $O_{ObVR}$ $O_I$ | | | |
| Teachers | | | Round 2 teachers | | | Round 2 teachers |
| Teaching Cases | R | | $O_{ST}$ X $O_S$ $O_{ST}$ | | | $O_{ST}$ |
| Looking at Student Work | R | | $O_{ST}$ X $O_S$ $O_{ST}$ | | | $O_{ST}$ |
| Content | R | | $O_{ST}$ X $O_S$ $O_{ST}$ | | | $O_{ST}$ |
| Control | R | | $O_{ST}$ — $O_{ST}$ | | X | - |
| Round 2 students | | | Round 2 student cohort | | | Follow-up student cohort |
| Teaching Cases | NR | | | $O_T$ $X_S$ $O_T$ | | $O_T$ $X_S$ $O_T$ |
| Looking at Student Work | NR | | $O_T$ $X_S$ $O_T$ | | | $O_T$ $X_S$ $O_T$ |
| Content | NR | | | $O_T$ $X_S$ $O_T$ | | $O_T$ $X_S$ $O_T$ |
| Control | NR | | | $O_T$ $X_S$ $O_T$ | | - |

R = randomly assigned
NR = not randomly assigned
$O_S$ = Teacher survey
$O_T$ = Content test
$O_I$ = Interview to assess PCK (intensive study only)
$O_{ObV}$ = Classroom observation and video (intensive study only)

$O_R$ = Teacher reflective interview while viewing video of own classroom (intensive study only)
X = *Understanding Electric Circuits* teacher course
$X_S$ = Student electric circuits unit
*Source* Developed by authors.

Each course was co-delivered by a pair of facilitators local to each research site, and each facilitator pair taught a different course in Round 1 than they did in Round 2 to avoid confounding facilitator and course effects. Having each facilitator pair teach multiple course variants introduces the possibility of contamination across conditions from their blurring distinctions between the experimental models. We controlled for such effects with a carefully counterbalanced design (see Table 7) such that (a) it included all possible sequences of two courses per facilitator pair over the two rounds; (b) there were overlapping assignments of facilitators so that each course was taught by more than one facilitator pair; and (c) each course variant was taught at least three times in each round. The linking of the data by overlapping assignments allows analysis of both facilitator and treatment effects, without confounding the two. Thus, the design controlled for facilitator main effects by having facilitators teach more than one course variant, and if there were order effects from facilitators having previously taught a different version of the course, they would be controlled through systematic variation in course sequences. This design does not eliminate contamination or order effects, but controls for them statistically.

Table 7. Counterbalanced Research Design with Three Intervention Models (A, B, and C) and a Control Group, as Originally Planned

| | | Round 1 | Round 2 | | |
|---|---|---|---|---|---|
| Site | Facilitator Pair | Fall 2007 | Winter 2007/08 | No. of intervention teachers | No. of control teachers |
| 1 | 1 | A | B | 20 | 7 |
| | 2 | B | C | 20 | 7 |
| 2 | 3 | C | A | 20 | 7 |
| | 4 | A | B | 20 | 7 |
| 3 | 5 | B | A | 20 | 7 |
| | 6 | C | B | 20 | 7 |
| 4 | 7 | C | A | 20 | 7 |
| | 8 | B | C | 20 | 7 |
| 5 | 9 | A | C | 20 | 7 |
| | 10 | B | A | 20 | 7 |
| 6 | 11 | A | C | 20 | 7 |
| | 12 | C | B | 20 | 7 |
| Totals | | | | 240 | 84 |

*Note*. A – Teaching Cases; B – Looking at Student Work; C – Content Immersion
Time distinctions are not relevant for the control group, so no D’s are shown in columns.
*Source* Developed by authors.

While the design was perfectly counterbalanced originally, school and district circumstances intervened and the design had to be modified (see Table 8). First, some sites turned out to be too small to support two concurrent courses over two rounds. This was addressed by increasing the number of sites to eight and conducting one course per round at half of them (sites 1-4). Another site, #3, discontinued participation after Round 1 as a result of a major change in leadership, but site 4 was able to compensate for this change by adding a third course. For various reasons, sites 4, 5, and 7 needed to teach some courses during the summers before and after Year 1; and, finally, one site (#7) needed to reverse the sequence of the courses for logistical reasons. Thus, while not perfect, the counterbalanced design was largely intact as actually implemented.

Table 8. Actual Counterbalanced Research Design with Three Intervention Models (A, B, and C) and a Control Group, at Random Assignment

| | | Round 1 | | Round 2 | | | | |
|---|---|---|---|---|---|---|---|---|
| Site | Facilitator pair | Summer 2007 | Fall 2007 | Winter 2007/08 | Summer 2008 | No. of intervention teachers | No. of control teachers | Total |
| 1 | 1 | | A | B | | 30 | 10 | 40 |
| 2 | 2 | | B | C | | 27 | 10 | 37 |
| 3 | 3 | | C | | | 15 | 10 | 25 |
| 4 | 4 | | A | B | A | 45 | 10 | 55 |
| 5 | 5 | | B | | A | 32 | 11 | 81 |
| | 6 | | C | B | | 27 | 11 | |
| 6 | 7 | | C | A | | 28 | 10 | 75 |
| | 8 | | B | C | | 26 | 11 | |
| 7 | 9 | A | | C | | 27 | 10 | 76 |
| | 10 | A | | B | | 29 | 10 | |
| 8 | 11 | | A | | C | 17 | 9 | 57 |
| | 12 | | C | B | | 21 | 10 | |
| Totals | | | | | | 324 | 122 | 446 |

*Note*. A – Teaching Cases; B – Looking at Student Work; C – Content Immersion
Time distinctions are not relevant for the control group, so no D's shown in columns.
Figures include only individuals with both pre- and post-instruction quiz data.
*Source* Authors' analysis of primary data collected for the study.

The three professional development courses were delivered eight times each during the first year of the study for a total of 24 times during the study (see Table 9), 12 in Round 1 in summer/fall 2007, and 12 in Round 2 in winter 2007/08. A total of 283 teachers participated in the first study events held: 201 intervention teachers in the professional development course, and 82 control teachers in project orientation and data collection meetings.

Table 9. Numbers of Professional Development Courses Implemented and Teacher Participants for Three Experimental Interventions and a Control Group in 2007-08

| Treatment | No. times offered in Round 1 | No. times offered in Round 2 | Total no. times offered | No. teachers |
|---|---|---|---|---|
| Teaching Cases | 5 | 3 | 8 | 70 |
| Looking at Student Work | 3 | 5 | 8 | 68 |
| Content Immersion | 4 | 4 | 8 | 63 |
| Control | - | - | - | 82 |
| Total | 12 | 12 | 24 | 283 |

*Source* Authors' analysis of primary data collected for the study.

### *Research Site Selection*

Regional research sites were identified through a series of discussions with district and county science educators in the United States. Initial contacts were made through an extensive network of WestEd contacts, and other contacts were identified in those conversations. The number of grade 4 teachers that were needed for the study restricted the search to urban districts with at least 10 middle schools, or to larger geographic regions consisting of many districts with a smaller number of middle schools per district. The features of regional science contexts that were criteria for participation included:

- Well-established, stable district or regional science program, so participants would not be teaching science for the first time.
- Strong science leadership (e.g., staff developers, teacher leaders, and district staff) from whom to draw local course facilitators, so as to test the courses' effects when delivered by professionals in the field (not the course developers). Availability of a qualified professional educator willing to serve as the local coordinator for the region.
- Standards-based curriculum for teaching science in place, along with necessary supporting resources for teachers and students, and variety in curricula across districts.
- Academically, culturally, and linguistically diverse student population.
- Proven ability to recruit teachers for professional development.
- Willingness to provide student test data.

The research sites that were established through this process included eight national research sites, four in the western United States (in Arizona, California, and Washington), and four in the eastern states (in Massachusetts, North Carolina, and Alabama). Each research site was either a large school district (three sites) or a collection of geographically close districts (four sites with four to eight districts each, and one with 15 districts), for a total of 39 districts in six states. Geographically close districts typically had a history of working together, with the support of a regional entity such as a regional subject area project, university, or county connection.

Site coordinators were hired as consultants to oversee study activities in each region, including recruiting teachers, arranging for meeting and course facilities, running local meetings at which they collected teacher test and survey data, pursuing missing teacher or student data as needed, and supporting local course facilitators and research staff with logistics as needed. Qualifications for serving as a site coordinator included extensive experience organizing and leading teacher professional development, strong local connections to teachers and district staff, and an orientation that was compatible with the Understanding Science professional development model, including a social constructivist perspective focusing on helping students and teachers learn about science through collaborative discourse. All but one of these individuals were employed in county offices of education, school districts, or at a local university as science educators,

### *Controlling Threats to Implementation Fidelity*

In addition to the contamination issues raised by facilitators teaching multiple course variants, there are other potential sources of contamination when different facilitators at the same site teach different versions of the same course. Most serious among them is spontaneous sharing of procedures and materials among the facilitators while they are teaching different variants. Rather than assuming that this does not occur, which is most likely to provide only the illusion of independence among interventions, our plan was to respectfully inform the facilitators of this issue to enlist their intentional cooperation.

The aim was to elicit the facilitators’ collaboration in maintaining the integrity of the models through building an understanding of, and commitment to, the research. To monitor and document the actual implementations, as well as differential advantages and problems that facilitators encountered in implementing the different models, phone conversations were held with facilitators after their sessions. This approach of teaching all facilitators about all three professional development models and carefully monitoring implementation was judged to be the best solution by both research design experts and professional development providers that we consulted.

As described above with regard to facilitator recruitment and training, facilitators were trained only on the model they were leading in Round 1 and efforts were made to surface and limit course-related bias. We also monitored implementations closely to ensure that enacted courses were as close as possible to the intended course features. Facilitator manuals for each variant clearly delineated and illustrated the features of each session. In phone debriefs with facilitators before their first session and between subsequent sessions, project staff provided reminders about the content, structure, and process of the upcoming sessions.

We also monitored implementations closely to ensure that enacted courses were as close as possible to the intended course features. Facilitator manuals for each variant clearly delineated and illustrated the features of each session. In phone debriefs with facilitators before their first session and between subsequent sessions, project staff provided reminders about the content, structure, and process of the upcoming sessions.

### *Recruitment of Teachers*

Because the topic of electric circuits is in elementary curricula primarily at the fourth grade, teachers at this grade were invited to participate. Statistical power estimates determined that a teacher sample of 256 (64 per experimental condition) would provide 80 percent or higher power to detect a minimum effect size of 0.20 (0.23 for English learners) at the student level and 0.51 at the teacher level (for Type I error = .05). Coordinators at each of the eight regional sites were asked to recruit a volunteer sample of up to 40 grade 4 teachers per site where only one course would be run in each round, and 80 at sites where two courses would be run in each round, recruitment targets that intentionally exceeded the number needed based on the power analyses, to allow for attrition. The number of teachers at any particular district in a region depended entirely on teacher interest, as participation in the study was voluntary. Participating teachers were recruited via email and announcements during professional meetings, and through existing teacher networks. Teachers were considered eligible to participate if they were currently teaching grade 4 in the 2007/08 school year, expected to be doing so again in the 2008/09 school year, and had not participated in previous Understanding Science courses. Teachers also had to consent to the study requirements, including to:

- Have at least one year of experience teaching fourth grade.
- Be randomly assigned to either an intervention group or the control group.
- If in the control group, attend two 2-hour project meetings at their site, one each in fall 2007 and winter/spring 2008.
- Attend a staff development course, *Understanding Electric Circuits,* in either summer 2007 (for intervention group teachers) or summer 2008 (for control group teachers).
- Teach and complete a classroom electric circuits unit by May 31, 2007.
- Provide survey and test data for the course evaluation.
- Provide teacher and student data again during a follow-up year if randomly selected to do so.

Participating teachers were not assumed to be representative of grade 4 teachers in their schools, districts, or states. It is intended that the results be generalized only to a sample of volunteer teachers. Moreover, resources did not permit systematic collection of data from non-participating grade 4 teachers in the same schools as teachers who did participate in the study. A sample of 446 teachers applied to participate in the study and met the selection criteria above. The teachers received a $650 stipend per participant, plus an additional stipend if they participated in an intensive classroom sub-study or in the follow-up data collection. Students were not randomly assigned, but rather were the students in all participating teachers' classes.

### *Random Assignment Procedure*

Teacher applicants were screened to ensure they met eligibility criteria, and were then randomly assigned to one of the experimental conditions. A teacher-level random assignment procedure was conducted that utilized the school as a blocking factor when there were two or more teacher participants per school, and utilized a constructed stratum of eight teachers as a blocking factor for teachers who were the only participants in their schools. A total of 446 teachers were randomly assigned to groups (324 to intervention and 122 to control) (see Table 10). The assignments were conducted by one of the research staff at Heller Research Associates.

Table 10. Numbers of Teacher Participants Recruited and Randomly Assigned to Intervention and Control Groups, by Research Site

| Site | Teaching Cases | Looking at Student Work | Content Immersion | Control group | Total |
|---|---|---|---|---|---|
| 1 | 15 | 15 | - | 10 | 40 |
| 2 | - | 14 | 13 | 10 | 37 |
| 3 | - | - | 15 | 10 | 25 |
| 4 | 30 | 15 | - | 10 | 55 |
| 5 | 18 | 28 | 13 | 22 | 81 |
| 6 | 13 | 13 | 28 | 21 | 75 |
| 7 | 28 | 15 | 13 | 20 | 76 |
| 8 | 11 | 7 | 20 | 19 | 57 |
| Full sample | 115 | 107 | 102 | 122 | 446 |

*Source* Authors' analysis of primary data collected for the study.

The recruitment process necessitated that a random assessment design that was both within-school and between-schools be used because, within each of the six research sites, there were two groups of teachers: (a) one group from schools with two or more participating teachers, and (b) another group from schools with only one participating teacher. For (a), we conducted the randomization within each school. For (b), all schools with only one teacher participant were ranked based on 2006 school-level state test scores, and then the ranked list was separated into strata consisting of eight teachers each (or fewer, for odd numbers of teachers). This procedure was followed within each regional site.

After these assignments, three of the large research sites were designated as intensive study sites, and a sub-sample of 36 teachers in Round 1 was randomly selected to participate in an intensive classroom study—12 from each of three sites (3 per experimental condition). These intensive study participants and a randomly selected half of the rest of the original sample of intervention teachers were then assigned to provide follow-up data one year later. (Teachers in the control group were not included in the follow-up study so they could receive the professional development course in summer 2008.)

### *Procedures to Minimize Contamination of Control Group Teachers*

One of the challenges of a design in which teachers are the unit of assignment within schools is that the close proximity of implementation group and control group teachers increases the possibility of control group contamination. In this study, there was a potential for control group teachers to learn about the content and approaches of the Understanding Science course, and even to look at the materials from the course; also, the implementation teachers could have spontaneously shared their newfound content knowledge or pedagogical strategies with their colleagues when they planned their electric circuits lessons.

Several steps were taken to prevent and correct breaches in the random assignment. The approach was to enlist the teachers' intentional cooperation in maintaining the integrity of the random assignments through building an understanding of, and commitment to, the research process. We asked all participants to sign both a consent agreement and a detailed Teacher Agreement to Protect the Study, both of which stipulated that they would (a) protect the differences among experimental conditions by not sharing or receiving course materials or information for the duration of the study, and (b) protect the validity of students' performance on the tests by not helping students answer the questions, and not looking at or copying the test questions.

### *Parent Consent Procedures*

The Institutional Review Board (IRB) required active parental consent to collect classroom video data. Moreover, many of the school districts participating in the study required active parental consent for pretest and posttest scores. Parental consent was solicited through a letter and consent form that was sent home with each student. The consent form described the purpose of the research and detailed the data for which we were requesting consent. A parent or guardian was asked to indicate whether he or she consented and sign the form, and the student carried the signed form back to his or her teacher.

### *Data Collection Procedures*

Key teacher outcomes included content knowledge in electric circuits, quality of explanations and applications of that content knowledge, pedagogical content knowledge, and pedagogical practices (see Table 11). Student outcomes were measured with instruments that capture student content knowledge in electric circuits and quality of explanations and applications of that content knowledge.

Table 11. Key Outcome Variables and Data Collection Measures, by Outcome Domain

| Variables | Measures |
|---|---|
| Teacher content knowledge in electric circuits | Science content test for teachers |
| Teacher explanations and applications in science | Open-ended questions on science content test |
| Teacher pedagogical content knowledge | Written questions on teacher content test<br>Teacher pedagogical content knowledge interview |
| Teacher practices | Observations of classroom lessons |
| Student content knowledge in electric circuits | Science content test for students |
| Student explanations and applications in science | Open-ended questions on science content test |

*Source* Authors' summary.

The data collection measures, samples, and procedures are summarized in Table 12. An online teaching background survey provided data on all teachers' professional experience and backgrounds in science teaching. As shown in Table 13, within each of the two data collection rounds in Year 1 and in the follow-up year, written teacher surveys on teaching practices and beliefs, as well as tests of electric circuits content knowledge were administered to all teachers. In Year 1 these measures were administered before any teachers in each round had taken one of the courses, and after the teachers had finished teaching about electric circuits during the school year.

Student content tests were administered before and after the electric circuits unit in classrooms of all participating teachers.

Table 12. Instruments, Samples, and Data Collection Procedures

| Instrument | Sample | Procedure |
|---|---|---|
| Written and online surveys | All facilitators and teachers; follow-up random sample of half the teachers per experimental condition | Packet sent to facilitators; all intervention teachers take pre- and post-course, during first and last course sessions and online. Control teachers complete in project meetings at each site and online. |
| Science content tests for teachers | All facilitators and teachers; follow-up random sample of half the teachers | Packet sent to facilitators; all intervention teachers take pre- and post-course, during first and last course sessions. Control teachers complete in project meetings at each site |
| Science content tests for students | Students of all teachers. | Teacher administers within two weeks before and two weeks after electric circuits unit.[a] |
| Teacher pedagogical content knowledge interview | Random sub-sample of nine per group. | Researcher administers pre and post interviews within two weeks before course, and after teacher has taught classroom unit. |
| Classroom videotaping and observation protocol | Two lessons in classrooms of random sub-sample of nine per group. | Videographer and observer collect data during two lessons in electric circuits unit. |
| Course session videotaping and observation | All course sessions at all sites. | Facilitators tape all sessions using single camera. Researchers observe and videotape Session 3 in each course in three intensive sites. |

[a] Sealed packets of student tests with standardized administration instructions and script sent to teachers. Teachers sign affidavit that they did not provide help to students other than reading test questions aloud.

*Source* Developed by authors.

Table 13. Data Collection Timing

| Method/Instrument | Round 1<br>(Sept & Dec 2007) | Round 2<br>(Jan & Apr 2008) | Follow-up<br>(Apr 2009) |
|---|---|---|---|
| Teacher surveys | ✓ | ✓ | ✓ |
| Teacher content tests | ✓ | ✓ | ✓ |
| Teacher pedagogical-content-knowledge interview | ✓ |  | ✓ |
| Classroom observation, audio- and videotaping | ✓ |  | ✓ |
| Student content tests | ✓ | ✓ | ✓ |
| Facilitator surveys | ✓ | ✓ |  |
| Facilitator content tests | ✓ | n/a |  |
| Facilitator post-session debriefing | ✓ | ✓ |  |
| Facilitator focus group interviews | ✓ | ✓ |  |
| Course session videotaping and observation | ✓ | ✓ |  |

*Source* Authors' summary.

As shown in Table 14, the subsample of intensive study teachers also participated in interviews designed to elicit pedagogical content knowledge before any teachers had taken one of the courses, and after the teachers had finished teaching about electric circuits. Electric circuits classroom lessons were videotaped on two days of the unit in a sample of classrooms.

Survey, content knowledge, and interview data were collected from the professional development facilitators as well. Every professional development course session at each site was videotaped by the course facilitators. In addition, research staff also videotaped and observed one session for each course taught in one of the three intensive sites, employing two cameras and enhanced sound, and completing supplemental field notes (see additional detail below under data collection instruments).

Data were collected in two rounds of professional development course implementation conducted from August-December 2007 and from January-June 2008.

### *Administration of student tests of electric circuits*

Prior to student test administration, packets of student tests were sent to participating teachers. These packets included instructions and an administration script, as well as a classroom information survey to be completed by the teacher about that class. Each student testing package included instructions about how to administer the tests, including rules on opening the tests, distributing the forms, collecting final documents, and securing completed answer sheets in sealed envelopes to be returned for data processing and scoring. Teachers administered the science tests to their own students, following a detailed testing protocol provided by the research team.

Table 14. Data Collection Locations

| Method/Instrument | All sites | Intensive study only |
|---|---|---|
| Teacher surveys | ✓ | |
| Teacher content tests | ✓ | |
| Teacher pedagogical-content-knowledge interview | | ✓ |
| Classroom observation, audio- and videotaping | | ✓ |
| Student content tests | ✓ | |
| Facilitator surveys | ✓ | |
| Facilitator content tests | ✓ | |
| Facilitator post-session debriefing | ✓ | |
| Facilitator focus group interviews | ✓ | |
| Professional development course session videotaping by facilitators | ✓ | |
| Professional development course session videotaping and observation by research staff | | ✓ |

*Source* Authors' summary.

Teachers administered the pretests during one period of class time in fall/winter 2007. Posttests were administered within two weeks after the completion of the class's electric circuits unit, whenever that occurred during the school year. Students who missed a test because they were absent were given a make-up test as soon as they returned to school.

The data process team applied quality assurance procedures to verify that the student data they received and stored in a database were accurate and secure. These procedures included matching of names, checking of test forms, comparison of student ID numbers and dates of birth on pretests and posttests, and verification of parental consent for each student.

### *Administration of teacher tests and surveys*

Intervention group teacher surveys and content tests were administered during the first and last sessions of each teacher course. Regional site coordinators administered surveys and tests to control group teachers in regional project meetings in fall 2007 and winter/spring 2008, after they completed teaching about electric circuits in their classes and students had taken their posttests. Site coordinators were provided with detailed test administration instructions to standardize procedures across research sites.

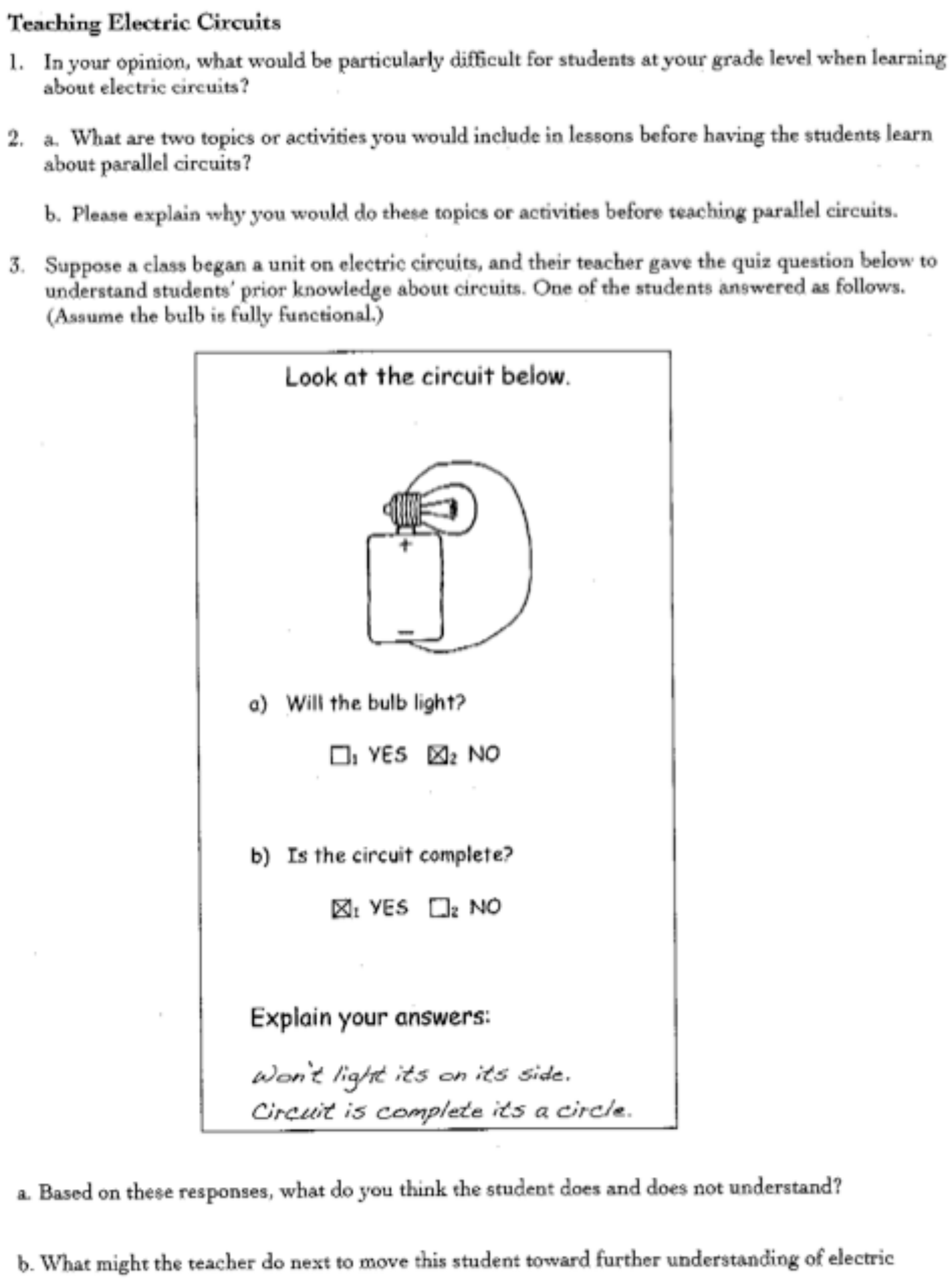
## ***Data Collection Instruments: Teacher, Classroom, and Student Outcomes***

***Teacher background and teaching context surveys.*** Teacher background surveys contained questions about (a) teacher background, including science training, teaching experience, and professional development experience; (b) class and school context, including school setting (urban, rural, etc., size of school) and student demographics; and (c) curriculum and textbook used to teach electric circuits. Two science teaching surveys (see Appendices x and y) administered at the beginning and end of Year 1 and end of Year 2 elicited a range of beliefs about science, science teaching, children's learning, etc., which could mediate the relationship between knowledge gain from the professional development and actual classroom practice. Post-professional development surveys given during the last session of each course included items to measure the degree and quality of the implementation of interventions (e.g., Cook et al., 1999; Foorman, Francis, Fletcher, Schatschneider, & Mehta, 1998), and teachers' ratings of the value and impact of the course.

***Science content test for teachers.*** A test that measures electric circuits was developed and validated for use in previous evaluations of the *Understanding Electric Circuits* course (Heller & Kaskowitz, 2004). This test was developed because, to our knowledge, there were and are no other validated tests on elementary grade electric circuits. The multiple-choice test was developed by creating a content framework specifying the targets of instruction by drawing on National Science Education Standards Benchmarks and FOSS and STC curricula, and creating test items that were aligned with the framework. Cronbach's alpha coefficient for the teacher test was determined to be .90. Teacher and student content tests were composed of slightly different questions, but covered the same material.

***Teacher explanations and applications in science.*** Three open-ended questions were included on the teacher content tests to serve as a measure of teachers' ability to explain and apply basic concepts in electric circuits. These questions, items 19, 20, and 26 on the test, asked (a) given a drawing of a battery, wire, and bulb, "explain why you think the circuit is or is not complete;" (b) given a drawing of a parallel circuit with one of its two bulbs missing, "explain why you think the bulb will or will not light;" and (c) given a drawing of two batteries and two bulbs, "draw wires to make a short circuit." Teachers' responses were scored using a rubric developed for each item.

***Teacher pedagogical content knowledge: Written questions.*** Open-ended questions were included on the teacher content tests to serve as a measure of teachers' pedagogical content knowledge for teaching electric circuits. The questions, shown in Figure 2, have been used in interview form in our previous studies, and were used in written form here for the first time. Specific aspects of pedagogical content knowledge that were target outcomes in this study were (a) knowledge about difficulties students have learning the content; (b) ability to hypothesize about science content understanding expressed in student work; and (c) pedagogical reasoning that incorporates strategies for building student understanding of specific science content required for task.



**Teaching Electric Circuits**

1. In your opinion, what would be particularly difficult for students at your grade level when learning about electric circuits?

2. a. What are two topics or activities you would include in lessons before having the students learn about parallel circuits?

   b. Please explain why you would do these topics or activities before teaching parallel circuits.

3. Suppose a class began a unit on electric circuits, and their teacher gave the quiz question below to understand students' prior knowledge about circuits. One of the students answered as follows. (Assume the bulb is fully functional.)

Look at the circuit below.

a) Will the bulb light?

☐1 YES ☒2 NO

b) Is the circuit complete?

☒1 YES ☐2 NO

Explain your answers:

Won't light its on its side.
Circuit is complete its a circle.

a. Based on these responses, what do you think the student does and does not understand?

b. What might the teacher do next to move this student toward further understanding of electric

Figure 2. Written open-ended questions measuring teacher pedagogical content knowledge.
*Source* Developed by authors.

***Teacher pedagogical content knowledge: Interviews.*** To obtain additional information about teachers' reasoning about teaching and learning of electric circuits, we used an in-depth interview that was developed and refined during previous studies (Heller, Daehler, Shinohara, & Kaskowitz, 2004). In the interview, teachers were asked to describe the kinds of conceptual difficulties students have understanding electric circuits, and how they would address those difficulties in the classroom. Teachers were shown samples of student work that exhibit a range of understanding of circuits, such as what happens when one bulb is unscrewed in a parallel circuit, and asked to describe what they would do to help each student better understand the science involved. This interview provides rich information about teachers' content knowledge, their understanding and anticipation of how different students think about and learn the science, and their own reasoning about how to teach particular concepts. A rubric for coding these interviews generates separate content and pedagogical content scores on an ordinal scale.

***Science content test for students.*** A student test on electric circuits was developed using the steps described for the teacher test. The test that was created is aligned with the *Understanding Science* project content framework, and includes questions that reflect the format and content of tasks in curricula such as FOSS and STC, and the TIMSS and NAEP assessments. Teachers provided demographic information about each student taking the test. Cronbach's alpha coefficients for the student test was determined to be .87.

***Student explanations and applications in science.*** Four open-ended questions were included on the student content tests to serve as a measure of students' ability to explain and apply basic concepts in electric circuits. These questions, items 12, 14, 15, and 22 on the test, asked (a) given a drawing of a battery, wire, and bulb, "explain why you think the bulb will or will not light; (b) given a drawing of a battery, wire, and bulb, "explain why you think the circuit is or is not complete;" (c) given a drawing of a parallel circuit with one of its two bulbs missing, "explain why you think the bulb will or will not light;" and (d) given a drawing of one battery and one bulb, "draw wires on the picture of the bulb and battery below to make a short circuit." Students' responses were scored using a rubric developed for each item.

***Classroom observations and videotaping.*** Two consecutive lessons for each intensive study teacher were observed and videotaped. Two methods were used to document instructional practices. First, we developed a Classroom Observation Protocol that both documents classroom practice and interactions in rich descriptive detail, and includes checklists and rating scales that generate ordinal scores along multiple dimensions. Second, we videotaped lessons in the unit on electric circuits with two tracks of audio, one from a lavalier on the teacher, and the other that recorded whole group and small group student interactions.

***Classroom interviews with teachers.*** Teachers were interviewed immediately before and after the two classroom observations. Interviews focused on the rationale for lesson design decisions, and on teachers' perceptions of what occurred during the classroom. The post-observation interviews included playing video clips of classroom moments. Interviews were structured to tap teachers' perception of the course content, design, and constituent activities. All interviews were audiotaped.

### *Data Collection Instruments: Professional Development Implementations*

Research under Goal 3 makes use of multiple data sources, including surveys, individual and focus group interviews, and extensive videotaping of professional development practice. We place particular emphasis on the affordances introduced by the video records. While classroom-based studies have increasingly employed elaborate video-based methods for capturing situated interactions among teachers and students, and for locating evidence of student learning, research on professional development has typically relied on interview- and survey-based reports (for recent exceptions, see Borko, Koellner, Jacobs, Roberts, Baldinger, & Risley, 2010; Kazemi & Franke, 2003; Kazemi et al., 2010). In doing so, it has provided limited insight into the construction of learning opportunity in professional development contexts, little appreciation for trajectories of teacher learning in those contexts, and scant attention to the role played by facilitators in supporting teacher learning. Although video-based research is resource- and labor-intensive, the interactions it captures achieve a degree of specificity and depth unlikely to be captured in observation protocols or field note summaries alone, and even more unlikely to be represented well in surveys or post-hoc interviews.

***Facilitator background survey***. Random assignment of facilitators to the sequence of experimental models in which they specialized should even out a large portion of the differences between groups in relation to background, experience, and demographics, but information about these variables was collected in a pre-study background survey so they can be described, statistically controlled, and studied in the analysis. The survey was designed to provide information on facilitators' experience in elementary science teaching and staff development, together with pre-training measures of facilitators' confidence in facilitating the Electric Circuits course. Facilitators completed the background survey prior to participating the Round 1 training. Data collected about the facilitators included: academic background and training in science; teaching credentials and degrees earned; years of teaching experience; years of experience teaching electric circuits at the elementary level; familiarity with elementary science curricula in electric circuits; experience as a learner in science professional development; training and experience as a provider of professional development in science; years of experience leading professional development in science; reported confidence in helping teachers understand twelve concepts encompassed by the electric circuits professional development (for example, the relationship between resistance and electric current); and their confidence in fifteen selected aspects of facilitation (for example, deciding how to respond to teachers' incomplete or incorrect science ideas).

***Facilitator content knowledge test***. A one-time administration of an electric circuits content knowledge test was used to determine that facilitators had a command of the knowledge sufficient to facilitate the Electric Circuits course. This instrument paralleled the content knowledge test used with teacher participants and made use of previously tested content items, together with open-ended items that elicited pedagogical content knowledge. Facilitators completed the content quiz on the last day of the facilitator training.

***Facilitator interview***. A one-on-one interview with individual facilitators, conducted on the penultimate day of the Round 1 training, focused on factors that are not easily tapped in a survey but that could help to explain observed facilitation of the Electric Circuits courses. The interview invited facilitators to describe their own goals and priorities as science staff developers, and to comment on what they anticipated both they and teachers would find easy or challenging about the Electric Circuits course.

***Facilitator post-training survey***. This brief survey, administered on the last day of the Round 1 facilitator training, served two purposes. First, it contained several items by which facilitators could assess the utility of the training in preparing them to lead the Electric Circuits courses.

Second, it included a post-training iteration of the facilitation confidence items included in the Facilitator Background Survey (pre-training). It also allowed for any further commentary on what facilitators anticipated would prove challenging in implementation.

***Videotape and observation records of facilitator training.*** Both rounds of facilitator training sessions were videotaped and all training materials collected. Supplemental audio-tapes were collected in small group activity. The training videotapes provide evidence of the intended enactment of each course model, the emphasis given to selected features of facilitation (for example, the use of visual representations to support conceptual understanding), and facilitators' initial responses to the course models and the approach to facilitation.

***Videotape and observation records of course sessions.*** Videotaping the course implementation served two purposes and employed two corresponding strategies. First, videotape records provided a means of maintaining the integrity of the separate interventions; we reasoned that the knowledge that they were being videotaped would heighten facilitators' sensitivity to preserving the intended components of each course model. In preparation for collecting this comprehensive record of course implementation, we supplied facilitation pairs with the equipment needed to create single-camera records of all course sessions, and prepared a 4-page guide to aid in securing a reasonable level of video quality. The resulting videotapes were logged and subsequently used to establish the design fidelity in each course.

Second, videotaping afforded the opportunity to investigate whether and how the facilitated professional development courses, as implemented, created the opportunity for teacher learning and to identify evidence of teacher learning. To ensure a smaller sample of high-quality video for that purpose, the research team observed and video-taped a smaller, purposively chosen sample of sessions in each of the three design configurations in the three intensive sites. Session 3, focused on the concept of electrical resistance, was considered an optimal choice for more intensive recording because it represented the point in the course at which (a) teachers would be making use of conceptual understandings developed in sessions 1 and 2 that most consistently map against the elementary curriculum; (b) teachers would be encountering new concepts that might not appear in the grade 4 curriculum but that were considered essential for deepening teachers' science knowledge for teaching; and (c) the group dynamics would most likely be well established. On these occasions, videotaping was handled more dynamically, with two cameras and enhanced sound-capturing capacity, to ensure records adequate for fine-grained analysis of interaction linked to the research questions (see Erickson, 1992; Hall, 2000).

***Facilitator debriefing interviews.*** Following sessions 1, 3 and 6 of each course, members of the research team conducted and audiotaped debriefing protocols with the facilitator pairs. The facilitator debriefing served both as a source of support for the facilitators and as a source of implementation data. It was designed to capture the facilitators' overall sense of how the course was proceeding, together with detailed accounts of what had gone well and what challenges had arisen in the context of specific activities.

***Facilitator post-course survey***. Following completion of each Electric Circuits course, the facilitators completed a brief survey in which they provided an assessment of the overall course success, the degree to which they believed key teacher outcomes were achieved, their experience of facilitation, and their advice regarding facilitator training.

***Facilitator focus group interviews***. The debriefing interviews and preliminary video analysis provided the basis for developing focus group interview questions and conducting facilitator focus group interviews. Focus group interviews for Round 1 were conducted in December 2007 at the initiation of the second round of facilitator training. Focus groups for Round 2 began in May 2008 and continued into early July to accommodate the staggered ending dates of Round 2 courses. The focus groups were all organized by intervention, but Round 2 groups included questions about the comparative strengths and limitations of the courses as perceived by the facilitators, as well as questions about the cumulative experience of facilitation across the two rounds.

### *Teacher and Student Analytic Samples*

The teacher analytic sample was defined as all teachers who were randomly assigned to each condition and who had valid posttest data. Just over a third of the 446 randomly assigned teachers (156) dropped out of the study before attending any project events or providing any data, generally because of scheduling conflicts or time constraints (see Table 15). The control group had the fewest teachers leaving the study at this stage, not surprisingly since they had much lower likelihood of a time conflict with brief project meetings than with 24-hour courses. Of the 290 teachers remaining, only 19 additional teachers (6.6 percent of those remaining at the start of Year 1) dropped out after attending one or more meetings or course sessions. The Content Immersion teachers had the highest dropout rate, with 12.5 percent leaving during Year 1.

Table 15. Number of Teachers Participating in Each Study Component

| | Group | | | | |
|---|---|---|---|---|---|
| Study component | Teaching Cases | Looking at Student Work | Content Immersion | Control | Total |
| Teachers randomly assigned | 115 | 107 | 102 | 122 | 446 |
| Dropped before first study event | 45 | 38 | 38 | 35 | 156 |
| Percent dropped before first event (percent) | 39.1 | 35.5 | 37.3 | 28.7 | 35.0 |
| Teachers at start of Year 1 | 70 | 69 | 64 | 87 | 290 |
| Attended first study event | 70 | 68 | 63 | 82 | 283 |
| Provided pretest data | 69 | 69 | 64 | 85 | 287 |
| Teacher Survey 1 | 68 | 69 | 64 | 84 | 285 |
| Teacher Quiz 1 | 68 | 69 | 64 | 84 | 285 |
| Provided posttest data | 69 | 66 | 56 | 80 | 271 |
| Teacher Survey 2 | 67 | 64 | 50 | 77 | 258 |
| Teacher Quiz 2 | 69 | 63 | 56 | 72 | 260 |
| Provided student data | 67 | 65 | 51 | 70 | 253 |
| Teachers at end of Year 1 | 69 | 66 | 56 | 80 | 271 |
| Attrition during Year 1 (percent) | 1.4 | 4.3 | 12.5 | 8.0 | 6.6 |

| Study component | Teaching Cases | Looking at Student Work | Content Immersion | Control | Total |
|---|---|---|---|---|---|
| | Group | | | | |
| Provided follow-up teacher data | 20 | 32 | 19 | n/a | 71 |
| Follow-up Teacher Survey 3 | 20 | 30 | 18 | n/a | 68 |
| Follow-up Teacher Quiz 3 | 20 | 30 | 18 | n/a | 68 |
| Provided follow-up student data | 20 | 30 | 18 | n/a | 68 |
| Teachers in follow-up (percent) | 28.6 | 46.4 | 29.7 | n/a | 35.3 |

*Source* Authors' analysis of primary data collected for the study.

A sample of 271 fourth grade teachers were retained in the study through the end of Year 1 and provided teacher data sets; 253 of these teachers provided pre- and posttest data from their students. Of the teachers retained in the study as of the end of Year 1, 71 also provided teacher and/or student data in the follow-up year. This sample corresponds to approximately 35 percent of the 290 teachers who started Year 1.

The 271 teachers retained in the study were distributed somewhat unevenly among the groups during each round of Year 1 and in the follow-up year (see Table 16). Experimental condition sample sizes in Year 1 ranged from 56 in the Content Immersion group to 80 in the control group, and in the follow-up year from 19 in the Content Immersion group to 32 in the Looking at Student Work group.

Table 16. Teacher and Student Sample Sizes by Study Round and Treatment

| Treatment | Teaching Cases | Looking at Student Work | Content Immersion | Control | Total |
|---|---|---|---|---|---|
| | | Teachers | | | |
| Round 1 | 41 | 27 | 32 | 42 | 142 |
| Round 2 | 28 | 39 | 24 | 38 | 129 |
| Total Year 1 | 69 | 66 | 56 | 80 | 271 |
| Follow-up | 20 | 32 | 19 | n/a | 71 |
| | | Students | | | |
| Round 1 | 855 | 595 | 593 | 670 | 2,713 |
| Round 2 | 608 | 819 | 490 | 818 | 2,735 |
| Total Year 1 | 1,463 | 1,414 | 1,083 | 1,488 | 5,448 |
| Follow-up | 403 | 609 | 372 | n/a | 1,384 |

*Source* Authors' analysis of primary data collected for the study.

The teacher analytic sample was defined as all teachers who were randomly assigned to each condition and who had valid posttest data.

### *Baseline Equivalence of Teacher Sample*

The internal validity of the study depends upon baseline equivalence among intervention group and control group teachers. Teacher-level characteristics were compared for the teachers who provided relevant data during Year 1.

Because participants were randomly assigned to groups, we expected teacher backgrounds to be similar across the four conditions. To verify this, we examined teachers' self-reported educational backgrounds, and teaching and professional development experience. The responses are summarized by intervention in Table 17. Overall, the backgrounds of the four groups appear comparable. Between 70 and 80 percent of the teacher sample had bachelor's degrees, and about half also had a master's degree. Because there were outliers in the distributions of teaching and professional development experience, median is a better measure of central tendency than mean. The median teaching experience was 7–9 years, with median electric circuits teaching experience of 3 years. Professional development experience in the past three years was also comparable, with medians of 20–24 hours in science and 6–8 hours in electric circuits. It is notable that the sample includes teachers with a wide range of teaching and professional development experience, spanning from novice to veteran in all groups.

Table 17. Educational, Teaching, and Professional Development Experience by Experimental Group

| Measure | Teaching Cases | Looking at Student Work | Content Immersion | Control |
|---|---|---|---|---|
| Bachelor's degree (percent) | 69.2 | 82.1 | 69.5 | 77.2 |
| Master's degree (percent) | 56.4 | 49.3 | 55.9 | 46.8 |
| Permanent teaching certification (percent) | 92.3 | 88.1 | 83.1 | 91.1 |
| Teaching experience (years) | | | | |
| Mean | 9.4 | 11.1 | 10.0 | 9.7 |
| Median | 7 | 9 | 7 | 8 |
| Range | 1–35 | 1–32 | 1–43 | 1–32 |
| *n* | 69 | 65 | 53 | 75 |
| Experience teaching electric circuits (years) | | | | |
| Mean | 4.0 | 4.1 | 3.6 | 4.3 |
| Median | 3 | 3 | 3 | 3 |
| Range | 0–13 | 1–25 | 0–13 | 0–18 |
| *n* | 62 | 49 | 38 | 53 |

| Measure | Teaching Cases | Looking at Student Work | Content Immersion | Control |
|---|---|---|---|---|
| Professional development in science in past 3 years (hours) | | | | |
| Mean | 55.6 | 63.7 | 48.4 | 46.4 |
| Median | 23 | 24 | 20 | 21 |
| Range | 0–648 | 2–800 | 2–400 | 2–300 |
| *n* | 43 | 31 | 28 | 42 |
| Professional development in electric circuits in past 3 years (hours) | | | | |
| Mean | 10.1 | 24.4 | 7.6 | 10.0 |
| Median | 6 | 8 | 6 | 6 |
| Range | 0–80 | 1–500 | 0–24 | 1–45 |
| *n* | 58 | 43 | 39 | 55 |

*Source* Authors' analysis of primary data collected for the study.

Teachers' demographics (Table 18) were also examined for equivalence. The teacher sample was approximately 80–90 percent female, about 60–70 percent White, about 9–14 percent Black, and 9–10 percent Hispanic or Latino. The control group had fewer Hispanic teachers and more White teachers than the other groups. However, in all categories there was variation among the four groups, but no consistent pattern indicated bias.

Table 18. Teacher Demographic Information, by Experimental Condition

| Characteristic | Teaching Cases | Looking at Student Work | Content Immersion | Control |
|---|---|---|---|---|
| *n* | 78 | 66 | 59 | 79 |
| Sex (percent) | | | | |
| Female | 87.2 | 83.3 | 78.0 | 87.3 |
| Male | 9.0 | 15.2 | 16.9 | 11.4 |
| Unknown or other | 3.8 | 1.5 | 5.1 | 1.3 |
| Race or national origin (percent) | | | | |
| White | 64.1 | 64.2 | 61.0 | 73.4 |
| Black | 14.1 | 10.4 | 8.5 | 8.9 |
| Hispanic | 10.3 | 9.0 | 8.5 | 1.3 |

| Characteristic | Teaching Cases | Looking at Student Work | Content Immersion | Control |
|---|---|---|---|---|
| Asian | 2.6 | 3.0 | 3.4 | 3.8 |
| Pacific Islander | 1.3 | 0.0 | 1.7 | 1.3 |
| More than one race | 2.6 | 1.5 | 3.4 | 3.8 |
| Other or unknown | 0.0 | 4.5 | 0.0 | 1.3 |

*Note.* White includes European; Black includes African American; Hispanic includes Latino or other Spanish origin; Asian includes Chinese, Indian, Japanese, Korean, and Vietnamese; and Pacific Islander includes Filipino, Guamanian or Chamorro, Native Hawaiian, Samoan, and other Pacific Islander.
*Source* Authors' analysis of primary data collected for the study.

Teachers' and students' pretest scores on the test of electric circuits were also examined for equivalence (see Table 19). All teacher means were between 56 and 61 percent, and student means between 47.5 and 48.9 percent, thus the experimental conditions started out with comparable baseline scores.

Table 19. Teacher and Student Background Information by Treatment

| Measure | Teaching Cases | Looking at Student Work | Content Immersion | Control | Total |
|---|---|---|---|---|---|
| Teacher pretest of electric circuits (percent correct) | | | | | |
| Mean | 60.9 | 56.3 | 57.7 | 56.5 | 57.8 |
| Standard deviation | 11.2 | 11.4 | 14.4 | 12.8 | 12.6 |
| *n* | 67 | 69 | 65 | 84 | 285 |
| Student pretest of electric circuits (percent correct) | | | | | |
| Mean | 47.7 | 47.5 | 47.8 | 48.9 | 48.0 |
| Standard deviation | 10.9 | 10.7 | 10.1 | 10.8 | 10.7 |
| *n* | 1,470 | 1,425 | 1,083 | 1,503 | 5,481 |

*Source* Authors' analysis of primary data collected for the study.

## Data Analysis Procedures

### *Goal 1: Investigate the relative impacts of the three professional development interventions on target teacher and student outcomes.*

The first three research questions focus on the treatment effects of the three professional development interventions on a number of target teacher and student outcomes (see Table 20). To address these questions about the relative impacts of the interventions, we evaluated the impact of each course on each outcome (see Table 20). Data for each outcome were first scored and then analyzed using hierarchical linear modeling analyses. One series of models evaluated teacher outcomes, and a second analyzed student outcomes. The relative value of the courses was also evaluated with respect to subgroups of special interest, including students of different race/ethnicity, English language proficiency, and sex. Outcomes for the first year of the study were evaluated separately from the follow-up data from the next year.

Table 20. Key Outcome Variables and Analysis Procedures, by Outcome Domain

| Variables | Analysis procedures |
|---|---|
| Teacher content knowledge in electric circuits | |
| Multiple-choice test items | Computed percent-correct scores on pretests and posttests. Fitted hierarchical linear models to gain-score data. |
| Open-ended test items | Hand-scored teacher written explanations and applications using task-specific rubric for each question. Fitted hierarchical linear models to gain scores from pretest to posttest. |
| Teacher pedagogical content knowledge | Hand-scored pretest and posttest open-ended pedagogical test item using task-specific rubrics. Fitted hierarchical linear models to posttest scores. |
| Teacher practices | Summarized structured checklists and rating scales from observations of classroom lessons. |
| Student content knowledge in electric circuits | |
| Multiple-choice test items | Computed percent-correct scores on pretests and posttests. Fitted hierarchical linear models to gain-score data. |
| Open-ended test items | Hand-scored student written explanations and applications using task-specific rubric for each question. Fitted hierarchical linear models to gain scores from content pretest to posttest. |

*Source* Authors' summary.

### *Scoring of Teacher and Student Content Knowledge Tests*

Two sets of scores were generated for both teachers and students as measures of their content knowledge about electric circuits: percent correct scores on multiple-choice test items, each of which received one point if correct and zero if incorrect; and human rater scores of written explanations and applications in response to open-ended test items. The written responses were scored using task-specific rubrics for each question (see appendices xx and yy for teacher and student rubrics, respectively).

### *Scoring of Teacher Pedagogical Content Knowledge*

Our conception of pedagogical content knowledge (PCK) centers on the core elements introduced in Shulman's original formulation (Shulman, 1986, 1987) as well as subsequent discussions of the PCK construct, and by empirical studies of PCK and its development. Defined most broadly, pedagogical content knowledge is the practical knowledge that enables teachers to transform the content and epistemology of a subject discipline for purposes of teaching. More specifically, the three core elements of our conception are:

- Teachers' capacity to *interpret subject disciplinary knowledge and practice for purposes of teaching* a school subject. Thus, teachers with a high level of PCK would be able to:
    - Explain the cognitive entailments of key concepts (what it would mean to understand them, what might make them difficult).

- Teachers' command of multiple, alternative, accurate and grade-appropriate *strategies for representing content knowledge*. In Shulman's terms, the teacher "finds multiple ways to represent the information as analogies, metaphors, examples, problems, demonstrations, and/or classroom activities" (1986, p. 9). Shulman also adds the ability to tailor teaching strategies to students—where grade level is one consideration but not the only one. Thus, teachers with a high level of PCK should be able to:
    - Communicate accurate science content knowledge to students, including descriptions of observable phenomena, demonstrations, and explanations of underlying mechanisms.
    - Incorporate in-depth knowledge of how to present the subject matter to learners using multiple representations of science phenomena and processes, such as electric current flow, circuit diagrams, written or verbal explanations, student hands-on activities, or student role-playing.
    - Make and justify instructional decisions in terms of intended learning outcomes.

- Teachers' *understanding of students' thinking and reasoning*; of what would constitute evidence of growing understanding or, alternatively, evidence of misconception or error (and related practices of formative assessment). Thus, a teacher with a high level of PCK would be able to:
    - Make explicit connections between specific student difficulties and instructional decisions.
    - Identify ways that conceptual difficulties are manifested in student work or behavior.

Our conception assumes that these elements of PCK interact with other elements of knowledge for teaching. Primary among these is foundational domain-specific *and* topic-specific content knowledge. Previous research also has shown that the development of pedagogical content knowledge depends upon strong content knowledge (Clermont, Krajcik, & Borko, 1993; Lederman, Gess-Newsome, & Latz, 1994; Smith & Neale, 1989).

These elements of PCK were incorporated into a coding scheme for analyzing teachers' responses to written questions involving interpreting samples of student work and describing instructional strategies they would use to address the specific difficulties embodied in those samples (see Table 22). The codes focused on: teacher actions (e.g., "I would explain to the student…." or "I would demonstrate…."), student actions (e.g, "I would have them build the circuit"), incorporation of instructional representations and activities to help students make sense of phenomena (e.g., "We would trace the flow of current through the wires and into the bulb"), and explication of the specific science learning goals that would be targeted. Each written response received between 0 and 2 points in each of the nine coding categories shown in Table 22.

The teachers' written pretest and posttest responses to the instructional strategies prompt (item 3b on the teacher content test) were analyzed here. The responses were transcribed into a spreadsheet, and presented in random order to a member of the research staff who was blind to any information about the teacher and whether the response was on the pretest or posttest. A second project staff member also scored a random sample of 15 percent of the full set of responses, corresponding to 30 responses, each receiving 9 scores, so 270 total score judgments. On this sample, the raters disagreed on only 6 scores, or 2.2 percent of the judgments.

Table 21. Scoring of Written Question, "3b. What might the teacher do next to move this student toward further understanding of electric circuits?"

| Category | Sample responses | Point value |
|---|---|---|
| Teacher actions | | |
| T1. Teacher explains, shows, or demonstrates, e.g., how/why a bulb will or will not light, how current flows, or short circuit. | I would show the student how the light bulb works.<br>Teach them about conductors and insulators. | 1 |
| Student actions | | |
| S1. Have student use materials to build the circuit to find ways to light bulb. | Have student create the circuit.<br>Provide the same materials from test question for the student to manipulate. | 1 |
| S2. Have student explain…draw or trace flow of current by pointing. | The teacher might put up multiple drawings of simple circuits with the bulbs in various locations and orientations (up or sideways). Have students trace path of electricity and highlight connection points on the bulb and battery. | 2 |
| S3. Have student draw or talk/write about what works and what doesn't work, or what lights what's complete. | Investigate and collect data by drawing what circuits will and will not light the bulb. Write a statement after comparing and contrasting the results above which explains the contributing factors that "lit" the bulb and didn't light the bulb. | 2 |
| S4. Have students act out model or metaphor for current. | Role play or further demonstrate "circuit". | 2 |
| S5. Discuss why | Discuss why this circuit won't light.<br>The teacher might have the student build this picture to find out if their predictions were accurate and then discuss the results. | 2 |

| Category | Sample responses | Point value |
|---|---|---|
| Science content | | |
| C1. Represent or examine architecture of light bulb, current through the bulb. (.5 if vague, e.g., "Look at how the light bulb functions.) | Show innards of bulb.<br>Students need to understand how a light bulb works. The point of entry must be different from the point of exit. Flow of current in a bulb goes base to filament to jacket or vice versa. | 2 |
| C2. Stress connections or contact points on bulb and battery. | Explain bottom of the bulb/sleeve are both needed to be connected by wire/battery to complete circuit and light the bulb. | 2 |
| C3. Focus on short circuit, warmth in circuit; or features of complete circuit | Ask, why is it getting hot? Do a lesson on short circuits.<br>Recording what a circuit must have to be complete would help further understanding | 2 |

*Note.* Total score is computed as the sum of points for all categories, minus one point if the response fits only category T1 or S1.
*Source* Developed by authors.

### *Impact Analyses*

The teacher model was a two-level HLM with teachers nested in professional development course groups, that is, the courses they took to learn the material. These course groups were expected to vary beyond differences related to the type of intervention, both as a result of the characteristics of the instructors and the interactions among the teachers within the class. The student model was a three-level HLM. Students were nested in classes taught by specific teachers nested in the professional development course groups.

Below are two equations, one for each of the models. The equations indicate what covariates and variables were involved in predicting gain scores. After fitting the model to the data, we then examined the coefficients that relate the strength and direction of these covariates with the outcomes.

The teacher model is:

$$Gain_{tp} = \alpha_0 + \alpha_1 TxA_p + \alpha_2 TxB_p + \alpha_3 TxC_p + \alpha_4 Nov_t + \alpha_5 Vet_t + \alpha_6 rnd2_p + \sum_{s=\{1,2,3,4,5,7,8\}} \alpha'_s Site_s + u_p^{(2)} + \varepsilon_{tp}$$

The formula for the basic model of student gains is:

$$Gain_{itp} = \beta_0 + \beta_1 TxA_p + \beta_2 TxB_p + \beta_3 TxC_p + \beta_4 Nov_t + \beta_5 Vet_t + \beta_6 rnd2_p + \sum_{s=\{1,2,3,4,5,7,8\}} \beta'_s Site_s + u_{tp}^{(2)} + u_p^{(3)} + \varepsilon_{itp}$$

Student gains (increase from pre- to posttest on the student circuits test) were regressed on teacher's experimental condition, research site, teacher's years of experience, and the round of the experiment as fixed effects. The model also included random intercepts for both the teacher and the individual professional development course. This allowed individual teachers' effectiveness to vary based on unmeasured covariates, as well as allowing the individual professional development courses to vary based on synergy of the course, effectiveness of the facilitators, and other unmeasured covariates. The terms are defined in Table 21.

Table 22. Definitions of Terms in Hierarchical Linear Models

| Term | Description |
|---|---|
| Gain | The difference in the circuit content knowledge test percent correct scores from pretest to posttest. Gain could theoretically range from -100 to 100. The same test with the same items was given at the pretest and posttest times. |
| $\alpha_0$, $\beta_0$ | The fixed intercepts for the two models (not interpretable in Models 2–3). |
| TxA, TxB, TxC | Dummy variables for interventions Teaching Cases, Looking at Student Work, and Content Immersion, taking the value 1 if the facilitators for the professional development group implemented the given intervention an 0 otherwise. Control is the reference group. |
| $u_p^{(3)}, u_p^{(2)}$ | The random intercept for the professional development group. These groups are defined as a group of teachers given a specific intervention together by a specific facilitator or facilitator pair. Note that this intercept is for variability *beyond* that due to the treatment effects. |
| $\text{Site}_k$ | A 0-1 dummy variable for whether the professional development group is from one of the Sites. Site 6 is the baseline site, so values of 0 for all other sites means the professional development course was taught in Site 6. We expect these coefficients to be negative, since Site 6 was taught by the creators of the material. |
| $u_{tp}^{(2)}$ | The random intercept for the impact of the teacher on the student's performance. |
| $\text{Novice}_t$, $\text{Veteran}_t$ | Novice is a dummy variable scored 1 if teacher *t* has taught 2 or fewer years. Veteran is a dummy variable scored 1 if teacher *t* has more than 8 years. If both are scored 0, that indicates teacher *t* has taught from 3-8 years. |

*Source* Authors' summary.

The superscripts of (2) and (3) indicate the levels of the model at which the random intercepts vary. The subscripts are *i* for individual student, *t* for teacher, and *p* for professional development group. Thus an individual student is indexed by their ID number, their teacher's ID number, and their teacher's professional development course's ID number. If a covariate has only a subset of the subscripts then we know that it is constant over the other subscripts. For instance $\text{TxA}_p$ denotes that interventions vary between professional development groups but not between teachers or students within the same professional development group.

The random intercepts are viewed as being drawn at random from a normal distribution with mean 0 and unknown standard deviation to be estimated from the data. In theory, a positive value for, say, $u^{(2)}_p$ , for some specific *p* would indicate that that group was somehow more effective in teaching the teachers the content knowledge measured by the posttest (possibly related to synergy of teachers or to those teachers being from a well-supported region).

All sites were fixed effects with Site 6 being the baseline site, and all other sites were compared to it. We believe that impact of the professional development courses taught during the second round of the experiment could potentially be larger than the first round due to the facilitators having had previous practice teaching the material. We did not, however, expect any variation in the control students or teachers from round to round, and this was verified by comparing gains, pretests, and posttests of all control students between Round 1 and Round 2. There were no discernable differences for any of these measures. Due to this, the round variable, rnd2, is a dummy variable with a value of 1 for students of teachers in one of the experimental conditions (A, B, or C) in the second round of the experiment. The rnd2 dummy variable is *not* set to 1 for round 2 control groups. This variable thus captures the mean effect of improved professional development course due to the facilitators' prior experience.

Veteran and Novice are dummy variables for the teachers of the students, with Veteran=1 if the teacher has taught for 8 or more years, and Novice=1 if the teacher has taught for 2 or fewer years. If both variables are 0, the teacher has taught from 3 to 8 years. These medium experience teachers are our baseline.

Note that the random effect of course is limited to only those courses for the interventions. The control groups, as they were not interacted with, were not expected to have any additional variation in effectiveness beyond that of teacher experience, site, and general teacher effect (as captured in the teacher random intercept). In particular, $u_p^{(3)} = 0$ for all controls.

### *Goal 2: Examine the relationships among gains in teacher content knowledge, teacher pedagogical content knowledge, and student achievement.*

> ***Question 4:*** Which, if either, is the better predictor of student achievement gains, teachers' post-course science content knowledge or pedagogical content knowledge?

We compared a series of HLM models to investigate the roles of teacher content knowledge and pedagogical content knowledge as potential mediators of treatment effects on student content test scores. One set of four student models were estimated in which experimental condition was left out, one with neither teacher content knowledge nor pedagogical content knowledge are included, one with only content knowledge, one with pedagogical content knowledge, and one with both. These indicated whether teacher content knowledge and/or pedagogical content knowledge were significant predictors of student achievement. A second set was also estimated with the same combinations but with experimental condition included. These will indicate how much of the variance in student outcomes are accounted for by teacher content knowledge and pedagogical content knowledge.

### *Goal 3: Examine the processes by which professional development designs achieve their effects on teacher knowledge and practice.*

Goal 3 entailed use of multiple data sources, including facilitator surveys, interviews, and content tests, but focused most heavily on the use of audio-visual recordings of the professional development course sessions to examine the two central research questions: How did the professional development courses engage teacher participation, achieve depth of talk about the key science ideas, elicit teachers' scientific reasoning, and enable focused about the teaching and learning of electric circuits content? What variations are evident in course implementation, and how do they relate to observed teacher and student knowledge outcomes?

Data from the surveys and interviews were used to confirm the capacity of newly-recruited facilitators to tackle the implementation demands associated with the professional development content and design. Results of the facilitator content quiz are presented are presented in the form of basic descriptive statistics (mean and SD scores by experimental condition). Results of on the confidence measures are reported by pre- and post mean scores, and results of the facilitators' assessment of the quality of facilitator training display both the mean scores and item distributions.

Comparative qualitative analyses of the audio-visual data took two forms during the period of project funding. Consistent with our interest in examining the relationship between professional development quality and subsequent teacher and student outcomes, we considered it potentially useful to construct a parsimonious index of course quality that could be incorporated in the HLM analysis. Toward that end, we constructed a rubric by which we could rate sampled course sessions with regard to content focus, depth, and accuracy; evidence of teacher reasoning about key science ideas; the extent and quality of teacher participation; and five dimensions of facilitation practice (support for transparent reasoning; contending with misconceptions and confusion; helping teachers to making connections across ideas and to achieve a general level of understanding; using materials, visual representations and inscriptions to support learning; and fostering teacher-to-teacher interaction). Focusing on a sample of target sessions (Session 3 in each course) across courses, we found variation along those dimensions at the extremes (especially with regard to facilitation practice and level of teacher participation), but overall the rubric proved too blunt an instrument to capture nuanced differences of the sort that have been exposed in more intensive case analyses. Reducing the 11 separate dimensions reflected on the rubric to a single "quality" score would have reduced meaningful variation still further. In effect, the predictive value of a global course quality index was deemed likely to be minimal, and the effort was abandoned. Thus, the only variable related to the professional development implementation incorporated in the HLM analysis was the measure of facilitator capacity reflected by facilitators' content test scores.

The principal, ongoing approach to comparative video analysis relies on intensive micro-analysis of purposively selected cases within the video corpus. The selected-case strategy took account of the highly labor-intensive nature of video analysis and was judged an appropriate use of staff resources. These case investigations employ conventions of interaction analysis to investigate the question of what constitutes opportunity for teacher learning in the professional development, and what might therefore help to explain teacher and student outcomes. Early singe-case investigations (Falk, 2009; Wong, 2010) have aided in working out codable dimensions of the professional development experience that inform cross-case comparisons across course configurations (interventions), between experienced facilitators and newly-trained facilitators, and between courses with higher and lower demonstrated outcomes.

Cross-case comparisons now underway have also capitalized on the early rubric development to establish a point of departure for a more fine-grained analysis of selected course sessions. The dimensions of the rubric provide the broad outlines for comparison, while the analysis employs methods of interaction analysis to examine precisely how opportunity for learning was co-constructed in the interaction among facilitators, teachers, and material resources. Because such analysis is labor-intensive, we developed a set of selection criteria to identify a subset of courses targeted for in-depth investigation. First, we sought sites that varied with regard to the most stringent outcome criterion, student learning gains. Second, we gave priority to sites designated as "intensive" sites, where the density of available data was highest (additional teacher interviews; classroom observation and video for a sample of teachers; additional course video). Finally, we conditioned course selection on video quality. The resulting pool of candidate sites included eleven courses, of which eight were from the designated intensive sites.

## Results for Goal 1: Relative Impacts of Interventions

The first goal of this study was to evaluate the relative impacts of the three professional development interventions on teacher and student outcomes. In this section we present the treatment effects on each target outcome.

### *Teacher Content Knowledge*

To address the first set of research questions about similarities and differences in the interventions' treatment effects, we estimated hierarchical linear models (HLMs) to evaluate the impact of each intervention on each outcome, beginning with teacher content knowledge. The analytic sample of teachers whose test scores were analyzed (shown in Table 23) included those with both pretests and posttests of electric circuits knowledge, as well as complete data on covariates that were included in the model. Analyses were conducted separately for teacher scores on the multiple-choice questions on the test, and for their written responses to open-ended questions.

Table 23. Numbers of Teacher Participants in Analysis of Year 1 Teacher Content Knowledge Test, by Experimental Group and Research Site

| Site | Teaching Cases | Looking at Student Work | Content Immersion | Control | Total |
|---|---|---|---|---|---|
| 1 | 5 | 7 | – | 6 | 18 |
| 2 | – | 9 | 6 | 9 | 24 |
| 3 | – | – | 9 | 5 | 14 |
| 4 | 23 | 11 | – | 8 | 42 |
| 5 | 9 | 18 | 7 | 16 | 50 |

| Site | Teaching Cases | Looking at Student Work | Content Immersion | Control | Total |
|---|---|---|---|---|---|
| 6 | 6 | 7 | 12 | 11 | 36 |
| 7 | 15 | 9 | 9 | 10 | 43 |
| 8 | 9 | 1 | 10 | 0 | 20 |
| Full sample | 67 | 62 | 53 | 65 | 247 |

*Note.* Analytic sample was defined as teachers with complete data, including pretest, posttest, and demographic/educational background covariates.
*Source* Authors' analysis of primary data collected for the study.

The three professional development interventions contained identical science investigations to strengthen teacher content knowledge about electric circuits. Results indicated that all three interventions caused large content test score gains for teachers. First, looking at the unadjusted pretest, posttest, and follow-up scores of the teachers by experimental condition (see Table 24), teachers in the three experimental conditions demonstrated mean gains on the test of electric circuits content knowledge of approximately 22 percentage points from Year 1 pretest to posttest, whereas control group teachers gained an average of 2.5 percentage points. On the posttest during the follow-up year, intervention teachers' mean gain scores of 20–23 percentage points remained considerably higher than the control teachers' gains in Year 1. Furthermore, as shown in Table 25, at every research site, intervention teachers achieved mean gain scores that were considerably higher than the mean gains for control teachers.

Table 24. Teachers' Unadjusted Mean Percent Correct and Gain Scores on Content Test, by Test and Experimental Condition

| Treatment | Pretest mean (SD) | Posttest mean (SD) | Follow-up mean (SD) | Year 1 mean gain (SD) | Year 1 to follow-up mean gain (SD) |
|---|---|---|---|---|---|
| Teaching Cases | 60.9<br>(11.2)<br>67 | 82.8<br>(10.8)<br>69 | 77.3<br>(11.2)<br>20 | 21.9<br>(11.0)<br>67 | 20.0<br>(11.2)<br>20 |
| Looking at Student Work | 56.3<br>(11.4)<br>69 | 77.8<br>(8.2)<br>62 | 75.4<br>(9.1)<br>29 | 21.5<br>(9.8)<br>62 | 22.8<br>(10.2)<br>29 |
| Content Immersion | 57.7<br>(14.4)<br>65 | 79.7<br>(11.8)<br>56 | 79.8<br>(11.5)<br>19 | 22.0<br>(12.6)<br>56 | 20.5<br>(12.8)<br>19 |
| Control | 56.5<br>(12.8)<br>84 | 59.0<br>(12.0)<br>72 | n/a | 2.5<br>(12.4)<br>72 | n/a |

Table 25. Teachers' Unadjusted Mean Percent Correct and Gain Scores on Year 1 Content Tests, by Test, Site, and Experimental Condition

| Site | Experimental condition | Measure | Pretest | Posttest | Gain |
|---|---|---|---|---|---|
| 1 | Interventions | *Mean* | 53.9 | 81.1 | 27.2 |
| | | *(SD)* | (10.5) | (8.8) | (9.7) |
| | | *n* | 12 | 12 | 12 |
| | Control | *Mean* | 56.7 | 58.3 | 1.7 |
| | | *(SD)* | (5.6) | (6.6) | (6.1) |
| | | *n* | 6 | 6 | 6 |
| 2 | Interventions | *Mean* | 55.9 | 77.3 | 21.5 |
| | | *(SD)* | (10.0) | (8.7) | (9.3) |
| | | *n* | 17 | 15 | 15 |
| | Control | *Mean* | 64.0 | 63.0 | -1.0 |
| | | *(SD)* | 13.8 | (12.5) | (13.1) |
| | | *n* | 10 | 10 | 10 |
| 3 | Interventions | *Mean* | 60.3 | 81.5 | 21.1 |
| | | *(SD)* | (8.4) | (11.6) | (10.0) |
| | | *n* | 10 | 9 | 9 |
| | Control | *Mean* | 60.6 | 57.8 | -2.8 |
| | | *(SD)* | (6.1) | (10.5) | (8.3) |
| | | *n* | 6 | 6 | 6 |
| 4 | Interventions | *Mean* | 59.4 | 81.1 | 21.7 |
| | | *(SD)* | (11.1) | (9.7) | (10.4) |
| | | *n* | 34 | 34 | 34 |
| | Control | *Mean* | 62.9 | 64.2 | 1.3 |
| | | *(SD)* | (6.3) | (10.9) | (8.6) |
| | | *n* | 8 | 8 | 8 |
| 5 | Interventions | *Mean* | 58.4 | 79.7 | 21.3 |
| | | *(SD)* | (10.6) | (8.3) | (9.4) |
| | | *n* | 36 | 36 | 36 |
| | Control | *Mean* | 47.8 | 52.0 | 4.3 |
| | | *(SD)* | (16.1) | (13.0) | (14.6) |
| | | *n* | 18 | 18 | 18 |
| 6 | Interventions | *Mean* | 53.1 | 77.8 | 24.7 |
| | | *(SD)* | (12.2) | (8.5) | (10.3) |
| | | *n* | 29 | 26 | 26 |
| | Control | *Mean* | 55.8 | 60.8 | 5.0 |
| | | *(SD)* | (10.6) | (9.9) | (10.2) |
| | | *n* | 12 | 12 | 12 |

| Site | Experimental condition | Measure | Pretest | Posttest | Gain |
|---|---|---|---|---|---|
| 7 | Interventions | *Mean* | 63.7 | 84.5 | 20.8 |
| | | *(SD)* | (15.2) | (12.8) | (14.0) |
| | | *n* | 34 | 34 | 34 |
| | Control | *Mean* | 59.7 | 61.4 | 1.6 |
| | | *(SD)* | (10.2) | (12.9) | (11.6) |
| | | *n* | 13 | 12 | 12 |
| 8 | Interventions | *Mean* | 58.4 | 81.9 | 23.5 |
| | | *(SD)* | (14.3) | (12.8) | (13.6) |
| | | *n* | 29 | 21 | 21 |
| | Control | *Mean* | 53.6 | Data lost | n/a |
| | | *(SD)* | (13.6) | | |
| | | *n* | 11 | | |

*Source* Authors' analysis of primary data collected for the study.

HLM analyses indicated that the gain scores of intervention teachers were significantly greater than those of teachers in the control group (see Table 26). On average, teachers in all three experimental conditions achieved adjusted gains of about 20 percentage points (effect sizes 1.9–2.1), whereas the control group's gain was less than 2.5 points. Furthermore, as shown in Table 27, the content test gains were maintained an additional year after the professional development, with adjusted gains ranging from 14–18 percentage points (effect sizes 1.1–1.5), still far greater than the control group gains in Year 1. No differences were found among the three courses in their impact on teacher content knowledge in the follow-up year.

Table 26. Impact Analysis of Teacher Adjusted Science Content Gain Scores in Percent Correct from Pretest to Posttest in Year 1

| | Adjusted Mean Gains[a] (Std deviation) | | | | | Unweighted | |
|---|---|---|---|---|---|---|---|
| Group | Intervention | Control | Difference (Std error) | $p$ | Confidence interval | Effect size | Total teacher sample size |
| Teaching Cases | 21.8<br>(10.3)<br>67 | 2.4<br>(9.9)<br>65 | 19.4***<br>(2.3) | .0001 | 14.9–24.0 | 2.0 | 247 |
| Looking at Student Work | 21.5<br>(10.3)<br>62 | 2.4<br>(9.9)<br>65 | 19.1***<br>(2.4) | .0001 | 14.4–23.8 | 1.9 | — |
| Content Immersion | 22.7<br>(12.9)<br>53 | 2.4<br>(9.9)<br>65 | 19.5***<br>(2.6) | .0001 | 15.3-25.3 | 2.1 | — |

*Significantly different from zero at the 0.05 level, two-tailed test. **Significantly different from zero at the 0.01 level, two-tailed test. ***Significantly different from zero at the 0.005 level, two-tailed test.

a. Model specification:

- Experimental condition (Teaching Cases, LASW, Content Immersion, control).
- Professional development course number (random intercept).

- Teacher pretest measure of content knowledge (pretest of electric circuits).
- Teacher teaching experience: ordinal, three-level scale of Novice (0–2 years), Intermediate (3–8 years), and Veteran (more than 8 years).
- Research site (1–8).
- Study round (1 or 2).

*Note.* Data were regression-adjusted using multilevel regression models to account for differences in baseline characteristics and study design characteristics. Effect sizes were calculated by dividing impact estimates by the unadjusted control-group standard deviation of the outcome variable.
*Source.* Authors' analysis of primary data collected for the study.

Table 27. Impact Analysis of Teacher Adjusted Science Content Gain Scores from Pretest in Year 1 to Posttest in Follow-Up Year

| | Adjusted Mean Gains[a] (Std deviation) | | | | | Unweighted | |
|---|---|---|---|---|---|---|---|
| Group | Intervention | Control (Year 1) | Difference (Std error) | *p* | Confidence interval | Effect size | Total teacher sample size |
| Teaching Cases | 14.0<br>(9.3)<br>20 | 2.4<br>(9.9)<br>65 | 10.9***<br>(3.5) | .002 | 4.1–17.7 | 1.1 | 134 |
| Looking at Student Work | 18.3<br>(11.5)<br>30 | 2.4<br>(9.9)<br>65 | 15.2***<br>(3.3) | .0001 | 8.7–21.7 | 1.5 | — |
| Content Immersion | 15.8<br>(12.3)<br>19 | 2.4<br>(9.9)<br>65 | 12.7***<br>(3.7) | .0001 | 5.5–20.0 | 1.3 | — |

*Significantly different from zero at the 0.05 level, two-tailed test. **Significantly different from zero at the 0.01 level, two-tailed test. ***Significantly different from zero at the 0.005 level, two-tailed test.
a. Model specification:

- Experimental condition (Teaching Cases, LASW, Content Immersion, control).
- Professional development course number (random intercept).
- Teacher pretest measure of content knowledge (pretest of electric circuits).
- Teacher teaching experience: ordinal, three-level scale of Novice (0–2 years), Intermediate (3–8 years), and Veteran (more than 8 years).
- Research site (1–8).
- Study round (1 or 2).

*Note.* Data were regression-adjusted using multilevel regression models to account for differences in baseline characteristics and study design characteristics. Effect sizes were calculated by dividing impact estimates by the unadjusted control-group standard deviation of the outcome variable.
*Source.* Authors' analysis of primary data collected for the study.

### *Teacher Science Explanations*

Teachers' responses to open-ended content test items followed the same pattern as for the multiple-choice items, with teachers in the three experimental conditions demonstrating higher mean gains (1–2 percentage points) from Year 1 pretest to posttest, whereas control group teachers gained an average of 0.1 percentage points (see Table 28). On the posttest during the follow-up year (Table 29), intervention teachers' mean gain scores of 0.6–1 percentage points remained considerably higher than the control teachers' gains of 0.1 in Year 1.

Table 28. Teacher Unadjusted Science Explanation Scores for Full Sample in Year 1

| Group | Measure | Year 1 Pretest | Year 1 Posttest | Year 1 Gain |
|---|---|---|---|---|
| Teaching Cases | *Mean* | 1.54 | 2.47 | 0.93 |
| | *(SD)* | 1.00 | 1.10 | |
| | *n* | 67 | 69 | |
| Looking at Student Work | *Mean* | 1.56 | 2.67 | 1.11 |
| | *(SD)* | 0.99 | 1.08 | |
| | *n* | 69 | 63 | |
| Content Immersion | *Mean* | 1.56 | 2.44 | 0.88 |
| | *(SD)* | 1.04 | 1.00 | |
| | *n* | 62 | 54 | |
| Control | *Mean* | 1.19 | 1.32 | 0.13 |
| | *(SD)* | 0.74 | 0.77 | |
| | *n* | 84 | 72 | |
| Total | *n* | 279 | 258 | |

*Source.* Authors' analysis of primary data collected for the study.

Table 29. Teacher Unadjusted Science Explanation Scores for Sample with Follow-Up Data

| Group | Measure | Year 1 Pretest | Year 1 Posttest | Year 1 Gain | Year 2 Follow-up | Year 2 Gain from pretest |
|---|---|---|---|---|---|---|
| Teaching Cases | Mean | 1.57 | 2.73 | 1.16 | 2.64 | 1.06 |
| | (SD) | 1.09 | 1.03 | | 0.92 | |
| | *n* | 20 | 20 | | 20 | |
| Looking at Student Work | Mean | 1.42 | 2.61 | 1.19 | 2.49 | 1.07 |
| | (SD) | 0.96 | 1.09 | | 1.07 | |
| | *n* | 30 | 30 | | 30 | |
| Content Immersion | Mean | 1.56 | 2.22 | 0.66 | 2.42 | 0.64 |
| | (SD) | 1.04 | 0.97 | | 0.89 | |
| | *n* | 19 | 19 | | 19 | |
| Control | Mean | 1.19 | 1.32 | 0.13 | – | – |
| | (SD) | 0.74 | 0.77 | | | |
| | *n* | 84 | 72 | | | |
| Total | *n* | 153 | 141 | | 69 | |

*Source.* Authors' analysis of primary data collected for the study.

HLM analyses indicated that the explanation gain scores of intervention teachers were significantly greater than those of teachers in the control group (see Table 30). On average, teachers in all three experimental conditions achieved adjusted gains of about 1 percentage point (effect sizes 0.9–1.1), whereas the control group's gain was less than 0.1 point. As shown in Table 31, only the Teaching Cases course led to gains that were maintained an additional year after the professional development, with an adjusted gain of 1.1 percentage points (effect size 1.2), still far greater than the control group gain of 0.2 in Year 1.

Table 30. Impact Analysis of Teacher Science Explanation Gain Scores from Pretest to Posttest in Year 1

| | Adjusted Mean Gains[a] (Std deviation) | | | | | Unweighted | |
|---|---|---|---|---|---|---|---|
| Group | Intervention | Control | Difference (Std error) | *p* | Confidence interval | Effect size | Total teacher sample size |
| Teaching Cases | 0.99<br>(1.2)<br>66 | 0.06<br>(0.9)<br>65 | 0.9***<br>(0.2) | .0001 | 0.5–1.4 | 1.0 | 246 |
| Looking at Student Work | 1.11<br>(1.4)<br>62 | 0.06<br>(0.9)<br>65 | 1.1***<br>(0.3) | .0001 | 0.6–1.5 | 1.1 | — |
| Content Immersion | 0.87<br>(1.4)<br>53 | 0.06<br>(0.9)<br>65 | 0.8***<br>(0.3) | .002 | 0.3–1.3 | 0.9 | — |

*Significantly different from zero at the 0.05 level, two-tailed test. **Significantly different from zero at the 0.01 level, two-tailed test. ***Significantly different from zero at the 0.005 level, two-tailed test.

a. Model specification:

- Experimental condition (Teaching Cases, LASW, Content Immersion, control).
- Professional development course number (random intercept).
- Teacher pretest measure of content knowledge (pretest of electric circuits).
- Teacher teaching experience: ordinal, three-level scale of Novice (0–2 years), Intermediate (3–8 years), and Veteran (more than 8 years).
- Research site (1–8).
- Study round (1 or 2).

*Note.* Data were regression-adjusted using multilevel regression models to account for differences in baseline characteristics and study design characteristics. Effect sizes were calculated by dividing impact estimates by the unadjusted control-group standard deviation of the outcome variable.

*Source.* Authors' analysis of primary data collected for the study.

Table 31. Impact Analysis of Teacher Adjusted Science Explanation Gain Scores from Pretest in Year 1 to Posttest in Follow-Up Year

| | Adjusted Mean Gains[a] (Std deviation) | | | | | Unweighted | |
|---|---|---|---|---|---|---|---|
| Group | Intervention | Control | Difference (Std error) | *p* | Confidence interval | Effect size | Total teacher sample size |
| Teaching Cases | 1.35<br>(1.0)<br>20 | 0.23<br>(0.9)<br>65 | 1.1***<br>(0.4) | .004 | 0.4–1.9 | 1.2 | 134 |
| Looking at Student Work | 0.88<br>(1.3)<br>30 | 0.23<br>(0.9)<br>65 | 0.7<br>(0.4) | .09 | -0.1–1.4 | 0.7 | — |
| Content Immersion | 0.36<br>(1.3)<br>19 | 0.23<br>(0.9)<br>65 | 0.1<br>(0.4) | .75 | -0.7–0.9 | 0.1 | — |

*Significantly different from zero at the 0.05 level, two-tailed test. **Significantly different from zero at the 0.01 level, two-tailed test. ***Significantly different from zero at the 0.005 level, two-tailed test.

a. Model specification:

- Experimental condition (Teaching Cases, LASW, Content Immersion, control).
- Professional development course number (random intercept).
- Teacher pretest measure of content knowledge (pretest of electric circuits).
- Teacher teaching experience: ordinal, three-level scale of Novice (0–2 years), Intermediate (3–8 years), and Veteran (more than 8 years).
- Research site (1–8).
- Study round (1 or 2).

*Note.* Data were regression-adjusted using multilevel regression models to account for differences in baseline characteristics and study design characteristics. Effect sizes were calculated by dividing impact estimates by the unadjusted control-group standard deviation of the outcome variable.

*Source.* Authors' analysis of primary data collected for the study.

### *Teacher Pedagogical Content Knowledge*

Teachers' written responses were elicited to a question about how they would address instructionally a student's limited understanding, as embodied in a student work sample. We present two views of the result, one that shows which codes were present in teachers' responses, and the second analyzing total PCK scores that were assigned to the responses.

Salient patterns in which codes were applied (see Table 32) were as follows:

*Teacher explains only*. This category corresponds to responses in which the teacher only referred to something he or she would explain or demonstrate, with no mention of any active student engagement. Interestingly, the highest proportion of Looking at Student Work teachers gave responses in this category—almost half of them only referred to something they would do, compared to about 30 percent of Content Immersion and control teachers, and only 20 percent of those who took the Teaching Cases course.

Table 32. Percent of Teachers Giving Each Category of Response to Written Pedagogical Content Knowledge Question, by Experimental Condition

| Response to item 3b, "What might the teacher do next to move this student toward further understanding of electric circuits?" | Teaching Cases | Looking at Student Work | Content Immersion | Control |
|---|---|---|---|---|
| *n* | 67 | 60 | 53 | 69 |
| Teacher explains only | 19.4 | 46.7 | 30.2 | 30.4 |
| Teacher has students do hands-on activity | | | | |
| Hands-on activity only | 10.4 | 8.3 | 13.2 | 39.1 |
| Hands-on activity as well as making meaning | 29.9 | 13.3 | 17.0 | 10.1 |
| Teacher engages students in making meaning | | | | |
| At least one strategy involving making meaning | 47.8 | 26.7 | 22.6 | 15.9 |
| Multiple strategies involving making meaning | 14.9 | 6.7 | 5.7 | 4.3 |
| Teacher specifies conceptual learning goal | | | | |
| Mention any key concept | 70.1 | 78.3 | 62.3 | 36.2 |
| Mention multiple key concepts | 19.4 | 25.0 | 18.9 | 7.2 |

*Source.* Authors' analysis of primary data collected for the study.

*Teacher has students do hands-on activity.* Two patterns emerged in relation to teacher responses that included having students work directly with bulbs, batteries, and wires. One set of responses included only mention of student hands-on work (for example, "I would have them build it."), with no reference to strategies for helping students make sense of what they observed. Close to 40 percent of the control teachers responded in this way, whereas fewer than 15 percent of any intervention group teachers did so. Only 8–10 percent of teachers who took the Teaching Cases and Looking at Student Work courses mentioned only hands-on activities. Both of these courses emphasized interpreting and building student understanding.

The second response pattern involved mention of student hands-on work, but with explicit reference to strategies for helping students make sense of what they observed. The largest proportion of Teaching Cases responded in this way, 30 percent, as compared with 10 percent of control teachers and 13–17 percent of teachers in the other two intervention groups.

*Teacher engages students in making meaning.* This category corresponds to mentions of strategies and representations that would engage students in understanding what they observe (such as by tracing the current through the circuit, or creating a T-chart to compare drawings of circuits that did and did not light the bulb). First, looking at mention of at least one such strategy, Teaching Cases again produced the highest proportion, close to 50 percent, whereas the other two intervention groups ranged from 23–27 percent, and only 16 percent of control teachers described meaning-making activities for students. Second, looking at responses that contained more than one such strategy, Teaching Cases again produced the highest proportion, 15 percent, whereas the other two intervention groups ranged from 6–7 percent, and only 4 percent of control teachers described multiple meaning-making activities for students.

*Teacher specifies conceptual learning goal.* The final category of response focused on mention of specific conceptual understanding of science content that the teacher wanted the students to understand. Here again we looked at how many teachers mentioned at least one conceptual learning goal, and here results ranged from the highest groups being 78 percent of Looking at Student Work teachers and 70 percent of Teaching Cases, to 62 percent of Content Immersion teachers. These proportions compared favorably to just over a third of the control teachers. Finally, looking at responses that included more than one key science learning goal, the Looking at Student Work group was again highest at 25 percent, with the other intervention groups at about 20 percent, compared to the considerably lower proportion of 7 percent of control teachers.

As described in the data analysis section, teachers' responses were also assigned point values. Unadjusted PCK scores show highest scores for Teaching Cases and Looking at Student Work (see Table 33) and lowest for control teachers. Results of HLM analyses (see Table 34) showed that Teaching Cases and Looking at Student Work courses produced significant increases in teacher pedagogical content knowledge compared to control scores (effect sizes 0.8–0.9), whereas the Content Immersion course did not.

Table 33. Teacher Unadjusted Pedagogical Content Knowledge Posttest Scores for Full Sample in Year 1

| Measure | Teaching Cases | Looking at Student Work | Content Immersion | Control |
|---|---|---|---|---|
| Mean | 4.04 | 3.82 | 3.21 | 2.33 |
| (SD) | 2.27 | 1.83 | 1.77 | 2.1 |
| *n* | 67 | 60 | 53 | 69 |

*Source.* Authors' analysis of primary data collected for the study.

Table 34. Impact Analysis of Teacher Pedagogical Content Knowledge Scores from Posttest in Year 1

| | Adjusted Mean Gains[a] (Std deviation) | | | | | Unweighted | |
|---|---|---|---|---|---|---|---|
| Group | Intervention | Control | Difference (Std error) | *p* | Confidence interval | Effect size | Total teacher sample size |
| Teaching Cases | 4.04<br>(2.3)<br>65 | 2.43<br>(1.9)<br>65 | 1.61***<br>(0.4) | .0001 | 0.8–2.4 | 0.87 | 239 |
| Looking at Student Work | 3.81<br>(1.8)<br>56 | 2.43<br>(1.9)<br>65 | 1.38***<br>(0.4) | .002 | 0.5–2.2 | 0.75 | — |
| Content Immersion | 3.24<br>(1.8)<br>53 | 2.43<br>(1.9)<br>65 | 0.81<br>(0.5) | .08 | -0.1–1.7 | 0.44 | — |

*Significantly different from zero at the 0.05 level, two-tailed test. **Significantly different from zero at the 0.01 level, two-tailed test. ***Significantly different from zero at the 0.005 level, two-tailed test.

a. Model specification:

- Experimental condition (Teaching Cases, LASW, Content Immersion, control).
- Professional development course number (random intercept).
- Teacher teaching experience: ordinal, three-level scale of Novice (0–2 years), Intermediate (3–8 years), and Veteran (more than 8 years).
- Research site (1–8).
- Study round (1 or 2).

*Note.* Data were regression-adjusted using multilevel regression models to account for differences in baseline characteristics and study design characteristics. Effect sizes were calculated by dividing impact estimates by the unadjusted control-group standard deviation of the outcome variable.

*Source.* Authors' analysis of primary data collected for the study.

### *Student Content Knowledge*

The ultimate question of interest is whether the teacher professional development interventions had an impact on student science achievement. As shown in Table 35, students of teachers in the three experimental conditions demonstrated mean gains of 19–22 percentage points from pretest to posttest on the test of electric circuits content knowledge, whereas students of control group teachers gained an average of 13 percentage points. Furthermore, as shown in Table 36, at every research site, students of intervention teachers achieved mean gain scores that appear higher than the overall mean gain for control teachers' students.

Table 35. Students' Unadjusted Mean Percent Correct and Gain Scores on Science Content Test, by Test and Experimental Condition

| | | Year 1 | | |
|---|---|---|---|---|
| Group | Measure | Pretest | Posttest | Gain |
| Teaching Cases | *Mean* | 47.5 | 66.3 | 18.9 |
| | *(SD)* | 3.8 | 8.6 | 8.3 |
| | Teacher *n* | 69 | 69 | |
| | Student *n* | 1,347 | 1,347 | |
| Looking at Student Work | *Mean* | 47.7 | 69.8 | 22.2 |
| | *(SD)* | 4.5 | 7.6 | 7.1 |
| | Teacher *n* | 69 | 69 | |
| | Student *n* | 1,392 | 1,392 | |
| Content Immersion | *Mean* | 47.7 | 67.3 | 19.6 |
| | *(SD)* | 3.8 | 7.3 | 6.6 |
| | Teacher *n* | 65 | 65 | |
| | Student *n* | 1,013 | 1,013 | |

| Group | Measure | Year 1<br>Pretest | <br>Posttest | <br>Gain |
|---|---|---|---|---|
| Control | *Mean* | 49.2 | 62.2 | 13.1 |
| | *(SD)* | 4.7 | 7.4 | 6.7 |
| | Teacher *n* | 87 | 87 | |
| | Student *n* | 1,494 | 1,494 | |
| Totals | Teacher *n* | 290 | 290 | |
| | Student *n* | 5,246 | 5,246 | |

*Note:* Classroom means are reported to eliminate effects of varying class sizes of different teachers.
*Source.* Authors' analysis of primary data collected for the study.

Table 36. Students' Unadjusted Mean Percent Correct and Gain Scores on Year 1 Content Tests, by Test, Site, and Experimental Condition

| Site | Experimental condition | Measure | Pretest | Posttest | Gain |
|---|---|---|---|---|---|
| 1 | Interventions | *Mean* | 48.1 | 71.1 | 22.9 |
| | | *(SD)* | (5.7) | (7.4) | (7.6) |
| | | *n* | 12 | 12 | 12 |
| | Control | *Mean* | 47.9 | 62.6 | 14.7 |
| | | *(SD)* | (3.3) | (9.7) | (6.7) |
| | | *n* | 6 | 6 | 6 |
| 2 | Interventions | *Mean* | 48.3 | 71.0 | 22.7 |
| | | *(SD)* | (3.9) | (8.6) | (7.6) |
| | | *n* | 17 | 17 | 17 |
| | Control | *Mean* | 49.0 | 62.1 | 13.1 |
| | | *(SD)* | 2.3 | (5.6) | (7.0) |
| | | *n* | 10 | 10 | 10 |
| 3 | Interventions | *Mean* | 47.7 | 65.1 | 17.4 |
| | | *(SD)* | (1.9) | (4.9) | (4.8) |
| | | *n* | 10 | 10 | 10 |
| | Control | *Mean* | 46.8 | 56.2 | 9.3 |
| | | *(SD)* | (2.6) | (4.4) | (2.8) |
| | | *n* | 6 | 6 | 6 |
| 4 | Interventions | *Mean* | 48.4 | 70.2 | 21.8 |
| | | *(SD)* | (2.5) | (6.3) | (6.7) |
| | | *n* | 34 | 34 | 34 |
| | Control | *Mean* | 47.8 | 66.4 | 18.6 |
| | | *(SD)* | (2.0) | (3.7) | (3.8) |
| | | *n* | 8 | 8 | 8 |

| Site | Experimental condition | Measure | Pretest | Posttest | Gain |
|---|---|---|---|---|---|
| 5 | Interventions | *Mean* | 45.7 | 65.9 | 20.1 |
| | | *(SD)* | (3.4) | (7.8) | (7.7) |
| | | *n* | 36 | 36 | 36 |
| | Control | *Mean* | 49.8 | 63.3 | 13.4 |
| | | *(SD)* | (6.3) | (7.5) | (7.8) |
| | | *n* | 18 | 18 | 18 |
| 6 | Interventions | *Mean* | 47.9 | 67.4 | 19.6 |
| | | *(SD)* | (5.0) | (7.9) | (6.2) |
| | | *n* | 30 | 30 | 30 |
| | Control | *Mean* | 49.4 | 62.6 | 13.1 |
| | | *(SD)* | (2.6) | (6.6) | (6.2) |
| | | *n* | 13 | 13 | 13 |
| 7 | Interventions | *Mean* | 47.6 | 68.4 | 20.7 |
| | | *(SD)* | (5.1) | (8.6) | (7.5) |
| | | *n* | 35 | 35 | 35 |
| | Control | *Mean* | 50.9 | 60.8 | 9.9 |
| | | *(SD)* | (6.5) | (9.0) | (7.5) |
| | | *n* | 15 | 15 | 15 |
| 8 | Interventions | *Mean* | 48.0 | 64.1 | 16.2 |
| | | *(SD)* | (3.4) | (9.0) | (9.5) |
| | | *n* | 29 | 29 | 29 |
| | Control | *Mean* | 48.6 | 60.3 | 11.7 |
| | | *(SD)* | (7.9) | (11.8) | (5.2) |
| | | *n* | 11 | 11 | 11 |

*Note:* Classroom means are reported to eliminate effects of varying class sizes of different teachers. Reported *n*'s are numbers of teachers.
*Source.* Authors' analysis of primary data collected for the study.

All three teacher courses had a significant positive effect (Table 37), producing higher student gains than those of students in control classes. The estimated difference in student gains of the intervention teachers compared to student gains of the control teachers represents the effect the intervention had on the outcome. For example, Teaching Cases has a predicted impact of 4.9 percentage points on student gain compared with control, meaning teachers who took that course would be expected to raise student scores 4.9 percentage points on the test more than comparable teachers who had not taken the course. The estimated effects for the other two courses appear greater, as they are both about 8 percentage points compared to 4.9 for Teaching Cases, and effect sizes for impact on student test scores were similarly higher, about 0.6 for Looking at Student Work and Content Immersion, versus 0.3 for Teaching Cases. However, Tukey's Honest Significant Difference pairwise comparisons among the four experimental conditions indicated that all three interventions were significantly greater than controls, but were not different from one another.

Table 37. Impact Analysis of Student Adjusted Science Content Gain Scores in Percent Correct from Pretest to Posttest in Year 1

| | Adjusted mean gains[a] (Std deviation) | | | | | Unweighted | |
|---|---|---|---|---|---|---|---|
| Group | Intervention | Control | Difference (Std error) | *p* | Confidence interval | Effect size | Student sample size (Teachers) |
| Teaching Cases | 18.4<br>(14.5) | 13.5<br>(14.3) | 4.9***<br>(1.4) | .0002 | 2.3–7.6 | 0.34 | 5,066<br>(238) |
| Looking at Student Work | 21.4<br>(15.1) | 13.5<br>(14.3) | 8.0***<br>(1.4) | .0001 | 5.3–10.7 | 0.56 | — |
| Content Immersion | 21.6<br>(14.6) | 13.5<br>(14.3) | 8.1***<br>(1.5) | .0001 | 5.2–11.0 | 0.57 | — |

*Significantly different from zero at the 0.05 level, two-tailed test. **Significantly different from zero at the 0.01 level, two-tailed test. ***Significantly different from zero at the 0.005 level, two-tailed test.

a. Model specification:

- Student demographic characteristics: sex (male, female), English learner status (beginning English learner, intermediate English learner, non-native fluent English proficient, native English speaker), and race/ethnicity (non-Hispanic White, Hispanic, Black, Native American, Asian, Pacific Islander, Middle Eastern/Arab, more than one, other).
- Student pretest measure of outcome variable.
- Teacher (random intercept).
- Professional development course number (random intercept).
- Teacher pretest measure of content knowledge (pretest of electric circuits).
- Teacher teaching experience: ordinal, three-level scale of Novice (0–2 years), Intermediate (3–8 years), and Veteran (more than 8 years).
- Experimental condition (Teaching Cases, LASW, Content Immersion, control).
- Research site (1–8).
- Study round (1 or 2).

*Note.* Data were regression-adjusted using multilevel regression models to account for differences in baseline characteristics and study design characteristics. Effect sizes were calculated by dividing impact estimates by the unadjusted control-group standard deviation of the outcome variable.

*Source.* Authors' analysis of primary data collected for the study.

Furthermore, the teachers' cohort of students in the school year after the study was completed demonstrated clear benefits of the teachers' professional development courses in the previous year (see Table 38). Follow-up students' adjusted content test gains ranged from 19 to 22 percentage points (effect sizes 0.4–0.7), still far greater than the control group gains of under 13 percentage points in Year 1. No differences were found among the three courses in their impact on student content knowledge, but all of them had powerful and sustained impact compared to the control condition.

Table 38. Impact Analysis of Student Adjusted Science Content Gain Scores from Pretest in Year 1 to Posttest in Follow-Up Year

| | Adjusted mean gains[a] (Std deviation) | | | | | Unweighted | |
|---|---|---|---|---|---|---|---|
| Group | Intervention | Control (from Year 1) | Difference (Std error) | *p* | Confidence interval | Effect size | Student sample size (Teachers) |
| Teaching Cases | 18.6<br>(14.5) | 12.6<br>(14.3) | 6.0***<br>(2.1) | .004 | 1.8–10.2 | 0.42 | 2,711<br>(131) |
| Looking at Student Work | 19.4<br>(15.1) | 12.6<br>(14.3) | 6.8***<br>(2.0) | .001 | 2.9–10.6 | 0.47 | — |
| Content Immersion | 22.4<br>(14.6) | 12.6<br>(14.3) | 9.8***<br>(2.3) | .0001 | 5.3–14.4 | 0.69 | — |

*Significantly different from zero at the 0.05 level, two-tailed test. **Significantly different from zero at the 0.01 level, two-tailed test. ***Significantly different from zero at the 0.005 level, two-tailed test.

a. Model specification:

- Student demographic characteristics: sex (male, female), English learner status (beginning English learner, intermediate English learner, non-native fluent English proficient, native English speaker), and race/ethnicity (non-Hispanic White, Hispanic, Black, Native American, Asian, Pacific Islander, Middle Eastern/Arab, more than one, other).
- Student pretest measure of outcome variable.
- Teacher (random intercept).
- Professional development course number (random intercept).
- Teacher pretest measure of content knowledge (pretest of electric circuits).
- Teacher teaching experience: ordinal, three-level scale of Novice (0–2 years), Intermediate (3–8 years), and Veteran (more than 8 years).
- Experimental condition (Teaching Cases, LASW, Content Immersion, control).
- Research site (1–8).
- Study round (1 or 2).

*Note.* Data were regression-adjusted using multilevel regression models to account for differences in baseline characteristics and study design characteristics. Effect sizes were calculated by dividing impact estimates by the unadjusted control-group standard deviation of the outcome variable.

*Source.* Authors' analysis of primary data collected for the study.

### *Student Science Explanations*

Students' responses to open-ended content test items did not follow the same pattern as for the multiple-choice items, in that students of teachers in the Looking at Student Work group achieved higher mean gains (0.17 percentage points) than controls from Year 1 pretest to posttest (see Tables 39 and 40). In the follow-up year (Table 41), however, students of both Teaching Cases and Looking at Student Work teachers demonstrated gain scores that were significantly higher than the control students' gains in Year 1.

Table 39. Student Unadjusted Science Explanation Scores for Teacher Sample with Follow-Up Data

| | | Year 1 | | | Year 2 | | |
|---|---|---|---|---|---|---|---|
| Group | Measure | Pretest | Posttest | Gain | Pretest | Posttest | Gain |
| Teaching Cases | Mean | 0.06 | 0.16 | 0.11 | 0.08 | 0.26 | 0.17 |
| | (SD) | 0.14 | 0.21 | | 0.13 | 0.23 | |
| | *n* | 1437 | 1439 | | 433 | 433 | |
| Looking at Student Work | Mean | 0.07 | 0.23 | 0.17 | 0.08 | 0.27 | 0.19 |
| | (SD) | 0.15 | 0.26 | | 0.13 | 0.26 | |
| | *n* | 1335 | 1336 | | 634 | 634 | |
| Content Immersion | Mean | 0.07 | 0.18 | 0.12 | 0.08 | 0.22 | 0.14 |
| | (SD) | 0.14 | 0.24 | | 0.12 | 0.22 | |
| | *n* | 1174 | 1174 | | 374 | 374 | |
| Control | Mean | 0.08 | 0.17 | 0.09 | — | — | — |
| | (SD) | 0.14 | 0.20 | | — | — | — |
| | *n* | 1383 | 1378 | | — | — | — |

*Source.* Authors' analysis of primary data collected for the study.

Table 40. Impact Analysis of Student Science Explanation Gain Scores from Pretest to Posttest in Year 1

| | Adjusted Mean Gains[a] (Std deviation) | | | | | Unweighted | |
|---|---|---|---|---|---|---|---|
| Group | Intervention | Control | Difference (Std error) | *p* | Confidence interval | Effect size | Total student sample size |
| Teaching Cases | 0.11<br>(0.21) | 0.11<br>(0.20) | 0.00<br>(0.02) | .91 | -0.05–0.04 | 0.00 | 1437 |
| Looking at Student Work | 0.17<br>(0.14) | 0.11<br>(0.20) | 0.06*<br>(0.02) | .01 | 0.01–0.10 | 0.32 | 1335 |
| Content Immersion | 0.12<br>(0.20) | 0.11<br>(0.20) | 0.01<br>(0.02) | .59 | -0.03–0.06 | 0.09 | 1174 |

*Significantly different from zero at the 0.05 level, two-tailed test. **Significantly different from zero at the 0.01 level, two-tailed test. ***Significantly different from zero at the 0.005 level, two-tailed test.

a. Model specification:

- Experimental condition (Teaching Cases, LASW, Content Immersion, control).
- Professional development course number (random intercept).
- Student demographic characteristics: sex (male, female), English learner status (beginning English learner, intermediate English learner, non-native fluent English proficient, native English speaker), and race/ethnicity (non-Hispanic White, Hispanic, Black, Native American, Asian, Pacific Islander, Middle Eastern/Arab, more than one, other).
- Student pretest measure of outcome variable.
- Teacher teaching experience: ordinal, three-level scale of Novice (0–2 years), Intermediate (3–8 years), and Veteran (more than 8 years).
- Research site (1–8).
- Study round (1 or 2).

*Note.* Data were regression-adjusted using multilevel regression models to account for differences in baseline characteristics and study design characteristics. Effect sizes were calculated by dividing impact estimates by the unadjusted control-group standard deviation of the outcome variable.
*Source.* Authors' analysis of primary data collected for the study.

Table 41. Impact Analysis of Student Adjusted Science Explanation Gain Scores from Pretest in Year 1 to Posttest in Follow-Up Year

| | Adjusted Mean Gains[a] (Std deviation) | | | | | Unweighted | |
|---|---|---|---|---|---|---|---|
| Group | Intervention | Control | Difference (Std error) | *p* | Confidence interval | Effect size | Total student sample size |
| Teaching Cases | 0.17<br>(0.17) | 0.09<br>(0.20) | 0.08**<br>(0.05) | .005 | 0.02–0.18 | 0.39 | 433 |
| Looking at Student Work | 0.18<br>(0.19) | 0.09<br>(0.20) | 0.09**<br>(0.04) | .006 | -0.01–0.17 | 0.36 | 634 |
| Content Immersion | 0.13<br>(0.17) | 0.09<br>(0.20) | 0.04<br>(0.05) | .19 | -0.06–0.14 | 0.11 | 374 |

*Significantly different from zero at the 0.05 level, two-tailed test. **Significantly different from zero at the 0.01 level, two-tailed test. ***Significantly different from zero at the 0.005 level, two-tailed test.
a. Model specification:

- Experimental condition (Teaching Cases, LASW, Content Immersion, control).
- Professional development course number (random intercept).
- Student demographic characteristics: sex (male, female), English learner status (beginning English learner, intermediate English learner, non-native fluent English proficient, native English speaker), and race/ethnicity (non-Hispanic White, Hispanic, Black, Native American, Asian, Pacific Islander, Middle Eastern/Arab, more than one, other).
- Student pretest measure of outcome variable.
- Teacher teaching experience: ordinal, three-level scale of Novice (0–2 years), Intermediate (3–8 years), and Veteran (more than 8 years).
- Research site (1–8).
- Study round (1 or 2).

*Note.* Data were regression-adjusted using multilevel regression models to account for differences in baseline characteristics and study design characteristics. Effect sizes were calculated by dividing impact estimates by the unadjusted control-group standard deviation of the outcome variable.
*Source.* Authors' analysis of primary data collected for the study.

### *Treatment Effects Related to Student Demographics*

One of the key intended outcomes for the professional development courses is to improve science achievement for diverse student populations, including typically underserved subgroups. The findings presented thus far have been for an analytic sample of all students in participating teachers' classrooms. For that full sample, students demonstrated mean gains of 19–22 percentage points from pre- to posttest on the test of electric circuits content knowledge, whereas students of control group teachers gained an average of 13 percentage points. Unadjusted mean gains for student subgroups classified at different levels of English language proficiency (see Table 42) show that the greater score increases for students of intervention teachers also occurred for all subgroups of English proficiency, for all three interventions tested.

There were no significant differences among student scores based either on sex or race/ethnicity.

Table 42. Student Unadjusted Science Content Gain Scores from Pretest to Posttest in Year 1, by Experimental Condition and English Language Proficiency Subgroup

| Sample | Measure | Teaching Cases | Looking at Student Work | Content Immersion | Control |
|---|---|---|---|---|---|
| English proficiency | | | | | |
| 1. Not English proficient—very little or no English (*n* = 106) | *Mean*<br>*(SD)* | 15.5<br>(16.5) | 10.7<br>(13.6) | 12.8<br>(10.5) | 6.0<br>(8.6) |
| 2. Intermediate English proficient—enough English to participate in classroom interactions (*n* = 444) | *Mean*<br>*(SD)* | 17.0<br>(14.4) | 19.9<br>(9.1) | 19.2<br>(10.9) | 9.2<br>(14.2) |
| 3. Fully English proficient—home or primary language not English (*n* = 644) | *Mean*<br>*(SD)* | 18.0<br>(10.6) | 21.8<br>(9.4) | 18.9<br>(11.1) | 13.4<br>(12.0) |
| 4. Fully English proficient—native English speaker (*n* = 3,980) | *Mean*<br>*(SD)* | 17.3<br>(10.1) | 21.7<br>(7.3) | 18.6<br>(7.6) | 11.6<br>(6.9) |
| Full sample (*n* = 5,174) | *Mean*<br>*(SD)* | 17.0<br>(12.9) | 18.5<br>(9.9) | 17.4<br>(10.0) | 10.1<br>(10.4) |

*Source.* Authors' analysis of primary data collected for the study.

### *Teacher Reports About Value of Courses*

Questions about the quality and impact of the professional development were included on an evaluation survey that was administered during the last session of each course implementation, as well as on the teacher post-instruction survey at the end of each year. Distributions of responses about the quality of the courses (see Table 22) indicated that close to 100 percent of the teachers rated the courses at least as highly as "better than the average professional development experience. However, teachers' overall opinions of the courses varied: 86 percent of Teaching Cases teachers rating it most highly, as "among the best of my professional development experiences", compared to 77 percent of the Looking at Student Work teachers and 66 percent of the Content Immersion teachers. When it comes to recommending the course to other teachers, however, over 95 percent of both Teaching Cases and Looking at Student Work teachers indicated they would definitely recommend the course, as did 88 percent of the Content Immersion teachers.

Table 43. Teacher Responses to Evaluation Survey Questions About Quality of Professional Development Courses

| Survey question | Teaching Cases | Looking at Student Work | Content Immersion |
|---|---|---|---|
| My overall opinion of this *Learning Science for Teaching* course: | | | |
| Percent responding "It was among the best of my professional development experiences." | 85.5 | 77.0 | 66.1 |
| Percent responding "It was better than the average professional development experience." | 13.0 | 23.0 | 30.4 |
| *n* | 59 | 47 | 37 |
| Would you recommend this course to other teachers? | | | |
| Percent responding "Definitely yes" | 95.7 | 95.1 | 87.5 |
| *n* | 66 | 58 | 49 |
| To what extent did the *Learning Science for Teaching* course influence what you included in your electric circuits unit? | | | |
| Year 1 post-instruction survey: | | | |
| Percent responding "A lot" | 88.5 | 95.3 | 87.2 |
| Percent responding "A little" | 11.5 | 4.7 | 12.8 |
| *n* | 69 | 61 | 56 |
| Follow-up post-instruction survey: | | | |
| Percent responding "A lot" | 100.0 | 96.7 | 89.5 |
| Percent responding "A little" | 0.0 | 3.3 | 10.5 |
| *n* | 20 | 30 | 19 |

*Source.* Authors' analysis of primary data collected for the study.

With respect to impact of each course on what the teachers included in their electric circuits units, at the end of Year 1, virtually all teachers responded that whichever course they took had at least some impact. The highest proportion of teachers in the Looking at Student Work course, 95 percent, responded "A lot" to this question, whereas just under 90 percent of teachers in the other two courses did so. The slightly higher proportion of Looking at Student Work teachers is most likely attributable to the task bank that teachers were given in the Looking at Student Work course, which they mentioned in interviews and open-ended questions as very valuable in their classes. By the end of the follow-up year, however, all but one of the teachers who had taken either the Teaching Cases or Looking at Student Work course indicated that the course had "A lot" of influence on what they included in their electric circuits unit. The proportion of Content Immersion teachers who reported the highest amount of influence on their teaching remained just under 90 percent.

Teachers were also asked their level of agreement with more specific questions about features of the courses. In order to describe the three groups' response patterns more generally, we summarized teachers' "Strongly agree" ratings (see Table 23) on each question, and computed average percents for four subclusters of survey questions that represent major aspects of the project's pedagogical content knowledge conceptual framework. These include questions explicitly related to *teaching*; about what is important for students to learn, or *learning goals*; about understanding *student thinking* about science ideas; and about the *science meanings* that comprise the content to be learned. Means for these clusters of items are provided at the bottom of Table 23. These means and the details for each item indicate that the Content Immersion course was rated positively by the lowest proportion of teachers in every category on the survey, and the other two courses were exactly comparable overall (an average of 77.7 percent of each group's teachers strongly agreeing with the items). Teaching Cases and Looking at Student Work were extremely close on the first three of these subclusters, and differed by only less than four percentage points on the science meanings subcluster. Interestingly, even though the Content Immersion course focused on developing teachers' metacognition about their own thinking and learning during the course sessions, less than 60 percent of the teachers who took this course reported gaining any insights into student thinking about the same content.

Table 44. Teacher Responses to Evaluation Survey Question About Features of Professional Development Courses

| Survey question | Category | Teaching Cases | Looking at Student Work | Content Immersion |
|---|---|---|---|---|
| Indicate how strongly you agree or disagree with the following statements (percent responding "Strongly agree"): (PPD7) | | | | |
| *n* | | 69 | 62 | 56 |
| i. This course probably will influence what I include in an electric circuits unit. | Teaching | 85.5 | 88.3 | 75.0 |
| j. This course probably will influence how I teach science. | Teaching | 82.4 | 86.9 | 69.6 |
| k. The course addressed the challenges of teaching electric circuits that have been most problematic for me. | Teaching | 76.5 | 65.0 | 53.7 |
| e. The course raised important teaching issues. | Teaching | 72.5 | 72.1 | 48.2 |
| a. The course covered concepts and phenomena that are essential to understanding and teaching electric circuits. | Learning goals | 88.4 | 88.1 | 83.9 |
| b. I gained a clearer understanding of what science ideas are important for students to learn. | Learning goals | 75.4 | 75.4 | 55.4 |
| d. The course helped me see what specifically makes some electric circuits concepts hard to understand. | Student thinking | 88.4 | 77.0 | 67.9 |
| f. The course gave me new ways to find out my students' ideas and thinking about science. | Student thinking | 73.5 | 77.0 | 52.7 |
| c. The course helped me see how different people think about electric circuits. | Student thinking | 67.6 | 72.1 | 58.9 |

| Survey question | Category | Teaching Cases | Looking at Student Work | Content Immersion |
|---|---|---|---|---|
| g. The course helped me understand meanings of specific scientific vocabulary and ways of using language. | Science meanings | 73.9 | 75.4 | 71.4 |
| h. The course helped me understand how to use precise language to communicate scientific ideas. | Science meanings | 71.0 | 77.0 | 57.1 |
|  | Teaching | 79.2 | 78.1 | 61.6 |
|  | Learning goals | 81.9 | 81.8 | 69.7 |
|  | Student thinking | 76.5 | 75.4 | 59.8 |
|  | Science meanings | 72.5 | 76.2 | 64.3 |
|  | Total Mean | 77.7 | 77.7 | 63.1 |
|  | Total *SD* | 7.2 | 7.4 | 11.1 |

*Source.* Authors' analysis of primary data collected for the study.

On particular questions, however, one or the other of the Teaching Cases or Looking at Student Work group was rated highly by a greater percentage of the course participants. For example, more than 10 percent more Teaching Cases than Looking at Student Work teachers responded "Strongly agree" on "k. The course addressed the challenges of teaching electric circuits that have been most problematic for me" (with Teaching Cases highest) and "d. The course helped me see what specifically makes some electric circuits concepts hard to understand." The statement with the strongest opposite pattern, of more Looking at Student Work teachers rating positively, was, "h. The course helped me understand how to use precise language to communicate scientific ideas," but no difference favoring the Looking at Student Work group over Teaching Cases reached 10 percent on any item. The statements for which the three groups were most similar were, "a. The course covered concepts and phenomena that are essential to understanding and teaching electric circuits," and "g. The course helped me understand meanings of specific scientific vocabulary and ways of using language." Conversely, statements for which the three groups differed most, with Content Immersion at least 20 percentage points below the other two courses, included, "e. The course raised important teaching issues," "f. The course gave me new ways to find out my students' ideas and thinking about science," and "b. I gained a clearer understanding of what science ideas are important for students to learn."

To provide additional insights into the teachers' experiences of the three courses, a different representation of the data in Table 23 is provided in Table 24. The latter contains the survey statements organized by the proportion of each group that strongly agreed with each: statements agreed to by 85 to 100 percent of the teachers, 75 to 84 percent, 60 to 74 percent, and 45 to 59 percent (no statements were strongly endorsed by less than 45 percent of any group). For each course, the statements agreed to by the highest proportion of teachers would most characterize that course, and conversely for the statements that least characterize the course.

Table 45. Statements about Professional Development Courses to Which Teachers in Each Group Responded “Strongly agree,” by Experimental Condition and Percent of Teachers

| Percent “Strongly agree” | Survey question |
|---|---|
| | Teaching Cases |
| 85-100 percent | a. The course covered concepts and phenomena that are essential to understanding and teaching electric circuits.<br>d. The course helped me see what specifically makes some electric circuits concepts hard to understand.<br>i. This course probably will influence what I include in an electric circuits unit. |
| 75-84 percent | j. This course probably will influence how I teach science.<br>k. The course addressed the challenges of teaching electric circuits that have been most problematic for me.<br>b. I gained a clearer understanding of what science ideas are important for students to learn. |
| 60-74 percent | g. The course helped me understand meanings of specific scientific vocabulary and ways of using language.<br>f. The course gave me new ways to find out my students’ ideas and thinking about science.<br>e. The course raised important teaching issues.<br>h. The course helped me understand how to use precise language to communicate scientific ideas.<br>c. The course helped me see how different people think about electric circuits. |
| 45-59 percent | None |
| | Looking at Student Work |
| 85-100 percent | i. This course probably will influence what I include in an electric circuits unit.<br>a. The course covered concepts and phenomena that are essential to understanding and teaching electric circuits.<br>j. This course probably will influence how I teach science. |
| 75-84 percent | d. The course helped me see what specifically makes some electric circuits concepts hard to understand.<br>f. The course gave me new ways to find out my students’ ideas and thinking about science.<br>h. The course helped me understand how to use precise language to communicate scientific ideas.<br>b. I gained a clearer understanding of what science ideas are important for students to learn.<br>g. The course helped me understand meanings of specific scientific vocabulary and ways of using language. |
| 60-74 percent | c. The course helped me see how different people think about electric circuits.<br>e. The course raised important teaching issues.<br>k. The course addressed the challenges of teaching electric circuits that have been most problematic for me. |
| 45-59 percent | None |

| Percent “Strongly agree” | Survey question |
|---|---|
| | Content Immersion |
| 85-100 percent | None |
| 75-84 percent | a. The course covered concepts and phenomena that are essential to understanding and teaching electric circuits.<br>i. This course probably will influence what I include in an electric circuits unit. |
| 60-74 percent | g. The course helped me understand meanings of specific scientific vocabulary and ways of using language.<br>j. This course probably will influence how I teach science.<br>d. The course helped me see what specifically makes some electric circuits concepts hard to understand. |
| 45-59 percent | c. The course helped me see how different people think about electric circuits.<br>h. The course helped me understand how to use precise language to communicate scientific ideas.<br>b. I gained a clearer understanding of what science ideas are important for students to learn.<br>k. The course addressed the challenges of teaching electric circuits that have been most problematic for me.<br>f. The course gave me new ways to find out my students’ ideas and thinking about science.<br>e. The course raised important teaching issues. |

*Source.* Authors’ analysis of primary data collected for the study.

### *Comparison of Effects of Three Interventions*

As summarized in Table 46, this randomized experimental study provided strong evidence of efficacy for the three professional development models tested in that all produced significant increases in teacher and student outcomes, but there were some important differences among their strengths. The three interventions all brought about highly significant gains in teachers’ and students’ scores on multiple-choice tests of science content knowledge, well beyond those of comparable control groups. The score increases for students of intervention teachers also occurred for all subgroups of English proficiency, from those speaking very little or no English to intermediate English learners, as well as native and non-native fluent-English-proficient students. There were no significant differences among student gains based either on sex or race/ethnicity. Furthermore, the powerful treatment effects were maintained in the school year following the study year, when both intervention teachers and their next cohort of students again showed gain scores significantly greater than those of controls.

Table 46. Effect Sizes and Significance Levels for Three Interventions Compared to Controls

| Measure | Teaching Cases | Looking at Student Work | Content Immersion |
|---|---|---|---|
| Teacher content knowledge Year 1 | 2.0*** | 1.9*** | 2.1*** |
| Teacher content knowledge follow-up year | 1.1*** | 1.5*** | 1.3*** |
| Teacher content explanations Year 1 | 1.0*** | 1.1*** | 0.9*** |
| Teacher content explanations follow-up year | 1.2*** | 0.7 | 0.1 |
| Teacher pedagogical content knowledge Year 1 | 0.9*** | 0.8*** | 0.4 |
| Student content knowledge Year 1 | 0.3*** | 0.6*** | 0.6*** |
| Student content knowledge follow-up year | 0.4*** | 0.5*** | 0.7*** |
| Student content explanations Year 1 | 0.0 | 0.3* | 0.1 |
| Student content explanations follow-up year | 0.4** | 0.4** | 0.1 |
| Teacher self-reports on surveys | | | |
| Percent of teachers responding “It was among the best of my professional development experiences.” | 85.5 | 77.0 | 66.1 |
| Percent responding “Strongly agree” to “The course addressed the challenges of teaching electric circuits that have been most problematic for me.” | 76.5 | 65.0 | 53.7 |
| Percent responding “Strongly agree” to “The course gave me new ways to find out my students’ ideas and thinking about science.” | 73.5 | 77.0 | 52.7 |

*Significantly different from zero at the 0.05 level, two-tailed test. **Significantly different from zero at the 0.01 level, two-tailed test. ***Significantly different from zero at the 0.005 level, two-tailed test.
*Source.* Authors’ analysis of primary data collected for the study.

With respect to the second measure of teacher and student content knowledge, the quality of explanations and applications in response to open-ended test items, a different pattern was observed (see Table 46). For teacher science explanations, in Year 1 of the study, all three interventions brought about highly significant gains in teachers’ scores as compared with controls. However, in the follow-up year, only the treatment effects of the Teaching Cases course were maintained and again produced significant increases. For students, in Year 1 only the Looking at Student Work course significantly raised science explanation scores. In the follow-up year, the delayed effects of the Teaching Cases course matched the effects of the Looking at Student Work course, and both significantly raised science explanation scores. Note that other than teachers’ Year 1 results, the Content Immersion course did not raise teachers’ or students’ scores on science explanations and applications.

The evidence of greater effectiveness for Teaching Cases and Looking at Student Work than for Content Immersion also manifested with respect to teacher pedagogical content knowledge (see Table 46). Only those two courses produced significant increases in teacher pedagogical content knowledge scores.

Finally, teachers' self-reports on surveys revealed consistently higher ratings for Teaching Cases and Looking at Student Work than for Content Immersion (see Table 46). In terms of overall quality, Teaching Cases was the most highly rated of the three, with over 85 percent of the teachers classifying it as "among the best of my professional development experiences." More than three quarters of the Teaching Cases teachers strongly agreed that the course, the only one of the three that engaged teachers in analysis of instructional practices, "addressed the challenges of teaching electric circuits that have been most problematic for me." The Looking at Student Work course, which focused teachers' attention on analyzing evidence in student work and in using classroom tasks that elicited student science understandings, was most highly rated for giving the teachers "new ways to find out my students' ideas and thinking about science."

## Results for Goal 2: Relationships Among Teacher and Student Outcomes

The findings thus far indicate that all three professional development courses increased both student and teacher science content knowledge, and that two of the courses (Teaching Cases and Looking at Student Work) increase teacher pedagogical content knowledge. The second goal of this study was to examine the relationships among gains in teacher content knowledge, teacher pedagogical content knowledge, and student achievement. In particular, the research question posed was which is the better predictor of student achievement gains, teachers' post-course science content knowledge or pedagogical content knowledge.

First, we determined whether teacher posttest content knowledge and/or pedagogical content knowledge predicted gains in student multiple-choice content test scores. We estimated the impacts with two HLM models, one each for teacher content knowledge and pedagogical content knowledge separately, with neither including experimental condition. Results indicated that both teacher outcomes were significant predictors of student test scores ($p < .001$ and $p < .05$ for teacher content knowledge and pedagogical content knowledge, respectively).

When both teacher content knowledge and pedagogical content knowledge were included in a model, content knowledge was a significant predictor of student performance on the multiple-choice test items ($t = 4.6$, $p < .001$), but pedagogical content knowledge was not. Interestingly, when student science explanation scores are modeled instead of the multiple-choice scores, both teacher content knowledge and pedagogical content knowledge are significant predictors of student performance ($p = .05$).

To determine whether teacher content knowledge accounts for most of the student outcomes, in which case it would be sufficient to bolster teachers' content knowledge in science as a means of producing student learning gains, we compared the results for the HLM model that had only teacher content knowledge, to the model that had both teacher content knowledge and the experimental condition dummy variables. A likelihood ratio test indicated that the models were indeed different ($p < 0.01$), and all three treatment effects were significantly positive ($p < 0.05$, $p < 0.005$, $p < 0.005$ for Teaching Cases, Looking at Student Work, and Content Immersion, respectively). The model coefficients indicate expected additional student gains *beyond those* *gains due to the teachers' content knowledge* of 3.4, 5.4, and 5.3 percentage points (effect sizes 0.26, 0.42, and 0.41, respectively). We conclude that all three teacher interventions do something to improve student test scores beyond that of merely improving teachers' content knowledge.

## Results for Goal 3: Professional Development Processes

Analysis linking professional development to teacher and student outcomes has proceeded in three stages, elaborated more fully below. First, we employed the facilitator survey and interview data to determine how well the newly-recruited facilitators were positioned to do the work of course facilitation; this analysis reflects our view that skillful facilitation was likely to be central to the effectiveness of the professional development. Second, we sampled selected sessions (session 3 across all courses in Round 1) from the full corpus of video to determine whether intervention integrity was being adequately maintained. Finally, we began the iterative and intensive process of substantive video analysis.

### *Analysis Stage 1: Determining Facilitators' Capacity to Implement the Professional Development Courses as Designed*

Although we have concentrated principally on delving into the professional development courses as they were implemented, a preliminary phase of analysis relied on facilitator survey and interview data to establish whether the newly-recruited facilitators commanded both the content knowledge and confidence needed for the work of facilitation. As the text and tables below indicate, the facilitators demonstrated a reasonable threshold of content knowledge and entered the first round of implementation with confidence in their ability to implement the courses as designed.

***Facilitators' teaching background.*** The newly recruited facilitators brought substantial teaching background to this work. The 20 facilitators ranged from 5 to 39 years of classroom experience (mean 19.3 years; median 20 years), and almost half (8) retained teaching responsibilities while serving as facilitators for the study. Their experience was heavily concentrated in elementary schools, and approximately two-thirds of them had taught at the fourth grade level (the target grade for this project). Every site had at least one facilitator with 4th grade experience and experience teaching electric circuits. The majority had taught units on electric circuits that encompass topics common to the elementary school curriculum: complete and incomplete circuits (14), conductors and insulators (13), series circuits (13), parallel circuits (12), and current flow (11); a few have taught resistance (4), short circuits (5), or path of least resistance (4), which are topics less often taught at the 4th grade. Half of the facilitators had experience with one or both of the most commonly used curriculum kits: FOSS (6); STC (4); or both (2).

***Facilitators' professional development background and responsibilities*.**[1] At the start of the project, the facilitators reported having a range of less than 1 year to 20 years of experience as

---

[1] As outlined above, facilitators were chosen by project staff and site coordinators according to the following criteria: 1) proven leadership in an existing professional development project, 2) understanding of science content, although not necessarily in electric circuits; 3) exceptional

professional developers (mean 6.3 years; median 4 years). In a pattern we believe to be characteristic of district-level elementary grades professional development, only 3 facilitators reported current responsibilities exclusively devoted to science professional development and/or instructional support in science. The remaining facilitators combined their professional development work in science with other activity. They continued as classroom teachers (8), conducted professional development in additional subjects that included literacy and math (6), and/or fulfilled various administrative responsibilities (4).

The facilitators' experience as professional developers in science tended to encompass approaches that could be characterized as "active engagement" (Garet et al., 2001) with science content and aspects of science teaching. For example, three-quarters of the facilitators (15) had led activities in which teachers completed hands-on science investigations, and nearly one-third (6) reported extensive involvement in such activity with teachers. More than half the facilitators had worked with teachers to examine examples of student work in science or on issues of student assessment. At the same time, facilitators made clear in both surveys and interviews that their prior professional development experience had not approached the depth of content understanding sought by the Learning Science for Teaching project. Facilitators commented that, by comparison with typical professional development in their districts, these courses would be "very in-depth" and "content-heavy." They anticipated that the level of content understanding sought by the professional development would prove challenging both for teachers and for themselves as facilitators, but also that the ambitious content goals would prove rewarding ("Some of the concepts are going to be a challenge, but that's the fun part of it.").

***Facilitators' level of content knowledge in electric circuits.*** The level of facilitators' content understanding is clearly an important factor in the kinds of facilitation they are able to provide and in the learning opportunity experienced by teachers. One explanation for teacher learning results would be the level of facilitators' own content understanding as measured and as represented in the professional development.

As measured by the content test administered at the end of the Round 1 training, all facilitators met a reasonable threshold of content knowledge prior to the first round of professional development implementation. Aggregate results show the mean percent correct to be 88.3 with range from 77.8 to 97.8 (see Tables 47 and 48). There was some minor variation by experimental condition, with Teaching Cases facilitators scoring somewhat higher overall (mean 90.6) than Looking at Student Work (mean 86.5) and Content Immersion (mean 87.9). All but one facilitation pair included at least one member with a higher-than-average score on the content quiz.

---

leadership and professional development skills, 4) knowledge of teaching issues specific to elementary science education; and 5) commitment to the proposed project.

Table 47. Number of Points on the Facilitator Quiz by Experimental Condition

| Measure | Teaching Cases | Looking at Student Work | Content Immersion | Total |
|---|---|---|---|---|
| Mean | 40.8 | 38.9 | 39.6 | 39.7 |
| Standard Deviation | 2.9 | 2.6 | 3.0 | 2.8 |
| Median | 41.3 | 38.5 | 40.5 | 40.0 |
| Minimum | 36.5 | 36.5 | 35.0 | 35.0 |
| Maximum | 44.0 | 42.5 | 43.0 | 44.0 |

*Note.* Maximum points possible are 45.
*Source.* Authors' analysis of primary data collected for the study.

Table 48. Percent Correct on the Facilitator Quiz by Experimental Condition

| Measure | Teaching Cases | Looking at Student Work | Content Immersion | Total |
|---|---|---|---|---|
| Mean | 90.6 | 86.5 | 87.9 | 88.3 |
| Standard Deviation | 6.5 | 5.8 | 6.7 | 6.2 |
| Median | 91.7 | 85.6 | 90.0 | 88.9 |
| Minimum | 81.1 | 81.1 | 77.8 | 77.8 |
| Maximum | 97.8 | 94.4 | 95.6 | 97.8 |

*Source.* Authors' analysis of primary data collected for the study.

***Facilitators' confidence and the reported contributions of facilitator training***. Although the facilitators had been recruited on the basis of relevant experience and expressed interest in the project, it remained the case that the facilitation demands embedded in the course were substantial. For that reason, we asked facilitators to report the degree of confidence they felt before and after their initial training regarding (a) their ability to help teachers learn specific concepts in electric circuits, and (b) their ability to employ specific kinds of facilitation practices.

*Confidence in facilitating content learning*: As displayed in Table 49, facilitators' confidence on selected electric circuits concepts prior to training varied widely, but was generally low to modest (means below 3.0 on 9 of 12 items). For example, facilitators were generally uncertain of their ability to help teachers understand the relationship between short circuits and path of least resistance (mean 2.2, or not very confident), but were more confident about their ability to help teachers grasp the difference between complete and incomplete circuits (mean 3.2, or somewhat confident). By the end of the training, facilitators' reported confidence on these items had strengthened considerably, with means on all items above 3.0.

Table 49. Facilitators’ Confidence in Helping Teachers Learn Selected Electric Circuits Concepts, Pre- and Post-Training (Round 1)

| Concept | Pre-training mean | Post-training mean | Difference |
|---|---|---|---|
| How to identify complete and incomplete circuits | 3.2 | 4.0 | 0.8 |
| How switches affect the flow of electric current | 3.2 | 3.8 | 0.6 |
| The relationship between resistance and electric current | 2.5 | 3.8 | 1.3 |
| How adding bulbs to a series circuit changes the brightness of the bulbs. | 3.0 | 4.0 | 1.0 |
| Factors that affect resistance | 2.3 | 3.6 | 1.3 |
| How current flows in series circuits | 2.8 | 3.9 | 1.1 |
| How current flows in a parallel circuits | 2.6 | 3.8 | 1.2 |
| The relationship between short circuits and the path of least resistance | 2.2 | 3.3 | 1.1 |
| Why a bulb may not shine even though electric current is flowing | 2.5 | 4.0 | 1.2 |
| The flow of electric current in terms of moving charges | 2.3 | 3.5 | 1.2 |
| What batteries do in series circuits | 2.3 | 3.9 | 1.6 |
| What batteries do in parallel circuits | 2.3 | 3.7 | 1.4 |

*Note.* Scale is 1 = not at all confident; 2 = not very confident; 3 =somewhat confident; 4 = confident.
*Source.* Authors’ analysis of primary data collected for the study.

*Confidence in using selected facilitation practices*: Table 50 displays facilitators’ generally high confidence in their ability to work with teachers in a professional development environment (13 of 15 items means > 3.0 prior to training), but indicates some uncertainty about selected practices that are central to the Learning Science for Teaching program, e.g., deciding how to respond to teachers’ incomplete or incorrect ideas.

By comparison to the confidence gains regarding the science content, the overall changes in Table 51 appear modest, although they generally show sustained or strengthened confidence. Based on additional data from the facilitator interviews, we have reason to believe that the group entered the project with the expectation that they would be successful working with teachers in a professional development environment; the two sets of pre-training scores might be explained as reflecting a certain distinction between content (which they knew to be specific) and professional development pedagogy (which they tended to treat in rather generic terms as “working with adults”). The post-training scores, together with interview comments and response to open-ended survey items, suggest that the facilitators gained a new appreciation for the sophistication of the facilitation practices and the ways in which those practices might specifically aid teachers’ science learning. It is not surprising, in that light, that the gains in facilitation confidence were modest overall. Indeed, it would have been possible to see facilitators’ confidence eroded during training had the facilitation expectations been experienced as too daunting. That did not occur. Table 54 displays facilitators’ overall favorable assessment of the contribution of the initial week-long facilitator training.

Table 50. Facilitators’ Reported Confidence in Using Selected Facilitation Practices, Pre- and Post-Training (Round 1)

| Practice | Pre-training mean | Post-training mean | Difference |
|---|---|---|---|
| Helping teachers gain an accurate understanding of key concepts in science | 3.2 | 3.70 | .50 |
| Eliciting teachers’ science reasoning during group discussions | 3.3 | 3.35 | .05 |
| Gauging the level of teacher understanding based on their discussion | 3.2 | 3.35 | .15 |
| Helping teachers feel comfortable expressing incomplete or incorrect science ideas | 3.4 | 3.75 | .35 |
| Deciding how to respond to teachers’ incomplete or incorrect science ideas | 3.1 | 3.15 | .05 |
| Using charts, diagrams, and other materials to develop and clarify teachers’ thinking | 3.5 | 3.65 | .15 |
| Deciding when to intervene to re-focus a group discussion | 3.3 | 3.20 | .10 |
| Deciding what to focus on in whole-group discussions | 3.4 | 3.30 | -.10 |
| Helping teachers anticipate what concepts will be difficult for students | 3.4 | 3.60 | .20 |
| Helping teachers identify patterns of common yet incorrect ideas in students’ science reasoning | 3.3 | 3.35 | .05 |
| Helping teachers see the logic in student reasoning even when students make errors | 3.1 | 3.40 | .30 |
| Facilitating discussions of written cases about electric circuits | 2.7 | 3.00 | .30 |
| Helping teachers select samples of student work that reveal interesting and important things about students’ science ideas | 2.9 | 3.00 | .10 |
| Helping teachers assess the merits and drawbacks of particular teaching choices (e.g., models, metaphors) | 3.1 | 3.40 | .30 |
| Helping teachers reflect on their own science learning process. | 3.4 | 3.50 | .10 |

*Note*. Scale is 1 = not at all confident; 2 = not very confident; 3 =somewhat confident; 4 = confident.
*Source.* Authors’ analysis of primary data collected for the study.

Table 51. Facilitators' Assessment of Initial Facilitator Training

| Survey statement | Strongly disagree | Disagree | Agree | Strongly agree | Mean |
|---|---|---|---|---|---|
| a. The training strengthened my own understanding of electric circuits. | 0 | 0 | 1 | 19 | 4.0 |
| b. The training strengthened my own understanding of concepts that are essential to teaching electric circuits. | 0 | 0 | 4 | 16 | 3.8 |
| c. The training helped me see what specifically makes some electric circuits concepts hard to understand. | 0 | 0 | 3 | 17 | 3.9 |
| d. The training gave me a clear understanding of how the Electric Circuits course is designed and sequenced. | 0 | 1 | 8 | 11 | 3.5 |
| e. The training showed me ways to elicit and work with teachers' ideas and thinking about electric circuits. | 0 | 0 | 3 | 17 | 3.9 |
| f. The training helped me anticipate what teachers might find challenging in the Electric Circuits course. | 0 | 0 | 3 | 17 | 3.9 |
| g. The training addressed the challenges of facilitating the Electric Circuits course that most concerned me | 0 | 0 | 7 | 13 | 3.7 |
| h. The training equipped me to make good use of the Facilitator Guide to plan and facilitate the course | 0 | 0 | 7 | 13 | 3.7 |
| i. Overall, I feel well prepared to start facilitating the Electric Circuits course. | 0 | 1 | 9 | 10 | 3.5 |

*Note.* One respondent marked an x between 2 and 3 on item i and noted the need for more studying of the content. The item was coded conservatively (2).
*Source.* Authors' analysis of primary data collected for the study.

### *Analysis Stage 2: Video-based Review of Fidelity to Course Design*

The first priority for data analysis was to assess fidelity to course design as reflected in the professional development course video records. Toward this end, we created a "fidelity log" template for each of the course configurations using the designed course segments/activities and suggested time allocations detailed in the Facilitation Guide. Given the large volume of video (nearly 500 hours of video across 8 sites and two rounds of facilitator training), we sampled course sessions 3 and 6 in each of the 8 sites and used an Excel spreadsheet to log the presence or absence of expected course segments, the elapsed time per segment, and notes regarding content emphasis and facilitator roles. This preliminary log served as a basis for judging intervention fidelity and also as a starting point for more in-depth analysis of the professional development practice and the learning opportunity created for teachers and facilitators. Although we did detect some variations in the time accorded to particular segments and activities, we found no evidence that any facilitator pair compromised the intervention design (for example, by introducing teaching case materials in a course designed to be content immersion only).

### *Analysis Stage 3: Video-based Analysis of Professional Development Course Implementation*

At the heart of this component of the research is the capacity for microanalysis of audiovisual recordings of the professional development courses, supplemented by data on facilitators' self-reported implementation experience derived from session-specific debriefing conversations, from the post-course surveys and focus groups, and from the videotaped facilitator training. It is to that analysis we turn to illuminate how the course design and the related material resources constitute opportunity for teacher learning, and whether and in what respects there prove to be consequential variations among the courses as implemented. Such microanalytic work is notoriously labor-intensive but potentially fruitful in fleshing out causal explanations (Maxwell, 2004) and in supplying guidance for future practice. As Erickson (1992) explains, "The microanalytic study of how interaction occurs is especially important when one wishes to reproduce an exemplary practice (e.g., the kind of classroom conversation in which students and teachers are excitedly engaged in reasoning together…) (p. 205).

To delve more fully into the character of the professional development experience, we organized the analysis around two main questions (identified above as questions 5 and 6):

- How did the professional development courses, as implemented, engage teacher participation, achieve depth of talk about the key science ideas, elicit teachers' scientific reasoning, and enable focused about the teaching and learning of electric circuits content?
- What variations are evident in course implementation, and how do they relate to observed teacher and student knowledge outcomes?

In the discussion below, we take those questions up in combination, with an emphasis on the contribution of the professional development to teachers' content understanding and the ways in which courses varied in supports for teacher content learning.

*Unpacking the contribution of professional development to teachers' content learning.* To investigate the contribution of the professional development to teachers' content learning, our analysis has focused on the Science Investigation segments in three courses that vary with regard to classroom outcomes and facilitator experience. Course 1, led by an expert facilitator, ranked highest among 24 courses with regard to student outcomes. Course 2, led by a newly-recruited facilitator, achieved results slightly below the experimental condition mean with regard to teacher knowledge, but above the experimental condition mean on student learning gains. Course 3, also led by a newly-recruited facilitator, was associated with lower results on both teacher knowledge and student learning outcomes. Both newly-recruited facilitators scored high on the content test (Facilitator 2, 94.4%; Facilitator 3, 95.6%).

We consider it important to situate the following analysis in the context of the overall findings. The newly trained facilitators, including those highlighted here in the courses we name Course 2 and Course 3, generally achieved a level of implementation sufficient to produce gains in teacher knowledge, classroom practice, and student learning. However, the implementation demands on facilitators were high, both in terms of their own content teaching knowledge and in terms of expected facilitation practices. Facilitators' prior experience had not required that they aspire to the depth of content understanding sought by this project. Most were unaccustomed to a professional development model that involved extensive supports for teachers' own sense-making and depended on facilitators' ability to elicit and work with teachers' incomplete, uncertain or incorrect understandings of science concepts. Analyses of the sort we summarize here focus on the extent to which the shifts toward intended practice were evident in the video records, and on aspects of facilitation that emerge as particularly difficult.

For purposes of this analysis, we rely on an analysis of the first of eight course sessions, in which the main content-related goal is to help teachers understand the nature of complete circuits, incomplete circuits, and short circuits. During the science investigation component of the session, teachers were first provided with a set of materials —a single battery, wire and light bulb —and worked in small groups to generate and document as many configurations as they could of circuit configurations in which the bulb lit or did not light. They then convened in a whole-group discussion to share and discuss their observations and to arrive at a generalized understanding of complete, incomplete and short circuits, and the relationship of those circuits to light as one indicator of a complete circuit. The analysis focuses on the whole-group discussion in the three courses.

In the space available, we concentrate on one principal way in which the qualitative analysis has yielded insight into the learning experience created through the interaction of facilitators, teachers, and the material resources of the courses. Specifically, we show how the visual representations and inscriptions embedded in the professional development curriculum functioned to support teacher learning; we also draw on a comparison across courses to show how variations in the strategic use of visual representations and inscriptions shaped the opportunity for teachers to reason in depth about the key science concepts and the relationships among them.

The analytic focus on visual representations and inscriptions derives in part from the design features of the professional development courses themselves; the Facilitator Guide[2] contains more than 130 visual representations and inscriptions (charts, diagrams, graphic organizers, lists, samples of student work) designed to supply cognitive supports for teacher reasoning and social supports for teacher communication. In addition, this analytic focus acknowledges the centrality of inscriptions and visual representations in science and in science teaching and learning more generally. Inscriptions such as graphs, charts, and diagrams are central to the construction of knowledge in scientific practice (Goodwin, 1994; Latour, 1990; Latour & Woolgar, 1986; Lynch, 1990; Stevens & Hall, 1998); much of scientific inquiry rests on the ability to use and interpret multiple forms of representation (Lemke, 1998; Hapgood, Magnusson, Palincsar, 2004). By focusing on inscriptions, the analysis takes account of a core feature of the intended professional development design and also bridges to a key feature of scientific practice and discourse.

In an elaborated case-study analysis of Course 1, Wong (2010) details three ways in which an expert facilitator's strategic use of representations and inscriptions engaged teachers in in-depth reasoning about key science concepts and supported teachers' progressive clarity about those

---

[2] See Wong (2010) for an analysis of the Facilitator Guide. The material resources for the course include the Participant Guide used by the teachers, but Wong's analysis focused solely on the Facilitator Guide, which offers additional guidance to facilitators about ways in which the learning supports embedded in the written curriculum might be exploited in practice.

concepts and their relationships to each other. In the discussion below, we begin by characterizing each of these strategic features of practice in Course 1 and then provide examples of the consequential variations evident in courses 2 and 3.

*Anchoring discussion to visual representations and inscriptions*. In Course 1, the facilitator consistently and systematically encouraged and prompted discussion that was heavily *grounded* in the use of inscriptions by (a) explicitly anchoring each of the session activities and discussions to key visual representations, (b) setting up the physical environment to allow all participants visual access to those key inscriptions, and (c) encouraging the participants to create or modify representations in such a way that it allowed the learners to operate on them as objects of thought.

Facilitators in Courses 2 and 3 also made use of the representations and inscriptions embedded in the curriculum, but made rather different choices regarding their selection, sequence, and functions. Those choices had consequences for the nature and depth of talk in which teachers engaged. For example, the facilitator of Course 3 opted not to use the T-Chart to organize teachers' circuit diagrams, thus turning the first two-thirds of the whole group discussion (18 of 27 minutes) into an exercise in categorizing the ten diagrams ("did it light?") and affording much less opportunity for analyzing the conceptual features of the diagrams. In a second example, the expert facilitator worked primarily with two forms of inscription: the T-chart (a graphic organizer), with its display of evidence; and the "What we're learning" chart (an evolving list). Together, these two inscriptions enabled teachers to work back and forth between assertions or generalizations and evidence. In Courses 2 and 3, facilitators devoted little or no time to discussing the particular features of the circuit diagrams (which represented the evidence brought by teachers from their small group observations), and substituted instead a "circuit connections chart" with six pre-drawn circuits of various types. Although the circuit connections chart lent itself (by design) to predicting circuit behavior, the facilitators in both cases used it to substitute for the circuit drawings as the basis for a discussion of complete and incomplete circuits. Finally, the facilitator in Course 1 made a special effort to ensure that all participants had visual and physical access to the inscriptions representations and inscriptions (supply dark markers for drawing circuit diagrams; waiting while teachers re-arranged chairs to provide clear sight lines and movement options). The facilitator in Course 2 similarly ensured that circuit diagrams were large enough and dark enough to be viewed easily by all participants. In Course 3, teachers' ability to work productively with the set of posted circuit diagrams was substantially diminished by the fact that they were not visible from a distance and had been posted in no particular order.

*Sustaining conceptual focus*. In Course 1, inscriptions served as tools for engaging learners in conceptual discussions, and for sustaining a focus on the key concepts encompassed in and specified for the session. While this might seem to be obvious and straightforward facilitation practice, prior research (e.g., Kesidou & Roseman, 2002) has shown that key science ideas are often lost amid a host of peripheral ideas and facts, or subordinated to an emphasis on process skills (for example, observational skills independent of key science ideas). In Course 1, the facilitator is shown to use inscriptions in the whole-group discussion to (a) invite extensive discussion about concepts, revealing aspects of the concepts that prove difficult or unfamiliar, (b) prioritize and highlight key concepts, (c) separate and distinguish ideas from one another, (e) document key conceptual questions, and (f) redirect off-topic conversations. As the teachers worked to analyze and annotate the diagrams displayed on the T-chart, they gradually arrived at specific assertions (recorded on the "what we're learning" chart) about the defining features of complete, incomplete, and short circuits. When they began to veer away from talk about the science concepts and into the terrain of classroom instruction, the facilitator used the two charts to re-direct conversation back to the concepts in play. In the closing episode of the Science Investigation, the facilitator engaged the teachers in creating an entirely new representation: collaboratively generated concept maps to convey their understanding of the key concepts and relationships. The concept maps, together with the teachers' verbal explanation of them, provide the facilitator a means of assessing group understanding at the close of the session.

Courses 2 and 3 also traverse the intended conceptual terrain of this first session. In both of these courses, teachers and facilitators focus on the concepts of complete and incomplete circuits. However, they do so in arguably less depth and with fewer opportunities for teachers to engage with one another and to build on one another's ideas. Indeed, it is instructive to compare the sheer incidence of teacher talk and the prevailing participation patterns across the three courses. In Course 1, teachers account for fully two-thirds of the recorded talk (as a percentage of transcript lines) and teacher-to-teacher exchanges account for more than half of all turns. In courses 2 and 3, teachers account for less than half the recorded talk (48% and 45% respectively) and teacher-to-teacher exchanges are sparse (11% and 8.5% respectively). Although the talk in both courses remains lively and engaged, it largely takes the form of dyadic exchanges between the facilitator and a succession of teachers.

The video records lend themselves to multiple points of comparison that illuminate how facilitators enable or constrain teachers' depth of reasoning about key concepts. For purposes of illustration, we focus here on the prompts employed by the facilitators as they make the transition from small group inquiry to whole group discussion. The prompts employed to launch the whole group activity shaped the discussion in particular ways. In Course 1, the facilitator's prompt specifically invited participants to construct a set of diagrams that reflected the observations conducted in small groups and that communicated a discovery: "So, we're going to switch over to whole-group in a minute and I'm going to ask people to start off by sharing either a circuit that *surprised* you, just like a drawing of it, or a pair of circuits that *helped you to learn something*."[3]

The prompt, with its emphasis on "surprise" or being "helped to learn something," lent itself to contributions that focused on unexpected discoveries, difficult concepts, or generative comparisons. The task — to come to agreement on one or two diagrams to display — similarly made it likely that the small groups would work together to identify the most central and difficult concepts. By inviting the participants to provide two drawings that aided learning, the facilitator encouraged the groups to focus on fruitful comparisons and develop a set of inscriptions that had

---

[3] A more typical prompt for whole-group discussions in classrooms and professional development sessions may take the form of an open-ended question such as "What did you learn?" In such cases, participants tend to produce a list of summary statements which are treated as known facts and could be considered as takeaways from the day's activities. Such summary statements, elicited at the outset of a whole group discussion, tend to mask areas of incomplete understanding within a group. When the individuals' contributions to a summary list are treated as correct, unproblematic, and understood by all, there is limited opportunity for the facilitator to assess the entire group's understanding.

the potential to support discussions about general rules. The resulting diagrams reflected both "surprises" (a circuit that is complete, producing heat, but does not light the bulb) and comparisons that clarified key conceptual elements (circuits that differed with regard to connection points but that both lit the bulb). In addition, by starting with the blank T-chart and inviting groups to post a group-generated diagram or pair of diagrams one group at a time, the facilitator remained in control of the pace at which the diagrams appeared for consideration and the kinds of discussion, annotation (added words, markings), and physical-re-arrangements of the diagrams by which teachers arrived at generalized claims linked to the evidence being displayed.

In courses 2 and 3, teachers heard recognizable variations on this prompt, but those variations oriented teachers differently to the task at hand and yielded different results. Teachers in Course 2 were asked to "Think of two pairs of circuits and put up two pairs, so go ahead and come up with two your two pairs, so four drawings that you did in your small groups and then go ahead and bring them up here and post them either on the lit side or did not light." While teachers Course 1 can be heard talking together what "surprised" them, teachers in Course 2 can be heard dividing up the labor of producing a set of lit and unlit circuit drawings ("I'll draw a lit one"). Both prompts had the effect of producing a data set on which teachers could subsequently focus discussion, but the prompt in Course 2 rendered the task as simply displaying visually (rather than analyzing or reflecting on) the results of the small group activity. Finally, teachers in Course 3 were asked to "draw a circuit that surprised you, maybe two from each group." In this prompt, the facilitator introduces the orienting notion of "surprise" and suggests that each group produce two such drawings, in principle preserving the possibility for productive comparisons. However, each teacher was then supplied with a piece of paper on which to draw and all teachers were asked to post individual drawings (all at once) on the front wall of the room. A prompt that might have spawned discussion about surprises that could then be tapped in the whole group discussion was thus converted to an individual task; further, the resulting data set was comprised almost entirely of "surprise" circuits of the same sort: complete circuits that did not light (8 of 11 posted circuits).

These three prompts to the identification, production and display of the circuit data set created different possibilities and challenges for the facilitators and the participants. The facilitator of Course 1 was positioned to help teachers consider a small set of drawings, one or two at a time, and to engage them in articulating the circuits' key distinguishing features and the patterns that marked them as similar and different. The facilitator of Course 2 was also aided by the organizing T-chart, enabling teachers to compare and contrast features of the lit and unlit circuits; however, the simultaneous posting of all drawings appeared to press in the direction of a fast-paced, serial "show-and-tell" as one person from each group ran quickly through a characterization of all the group's drawings at once. Finally, the facilitator and the teachers in Course 3 were placed in something of a bind by the absence of an organizing T-chart and the simultaneous posting of all the drawings; in combination, these conditions required additional work by the group to categorize the drawings as lit or unlit. Although many of the posted circuits were equivalent, each person's drawing was granted attention, perhaps reflecting a certain conventional classroom etiquette of attending individually to each person's work once it has been publicly displayed.

In the three selected courses, apparently subtle differences in orienting prompts resulted in different material conditions for subsequent thought and discussion (differences in the type of data available for consideration and in the way it was displayed for analysis); the prompts also differed in the normative conditions they established or reinforced with respect to teachers' interaction with one another and with the facilitator. The data prompts regarding the representation and display of observational data serve as one example of the ways in which visual representations and inscriptions potentially help to create and sustain conceptual focus, and the ways in which courses vary in their productive use. Other variations evident in the cross-course analysis and plausibly linked to differences in course outcomes include differences in the use of inscriptions to: identify and resolve conceptual confusion and uncertainty; distinguish concepts from one another; re-direct conversation as needed to sustain a focus on key concepts; and represent an understanding of the relationships among the key concepts for the session (in this instance, the relationship between complete and incomplete circuits and lit and unlit bulbs).

*Engaging in scientific reasoning practices*. In Course 1, inscriptions served as a vehicle for conceptual understanding by engaging the teachers collectively in *scientific reasoning practices*. Throughout the Science Investigation segment, participants were engaged in documenting and reporting observations, identifying and analyzing patterns in a shared data set, making generalized claims, supporting claims with evidence, and describing and representing the relationships between concepts.

The teacher learners in Course 1 (as in the other two courses) were asked first to represent the observations they had made in the prior small group activity (through drawing). By asking the group to share these observations about whether certain circuit configurations lit or did not light, the facilitator helped the group develop a pool of observations (a shared data set) from which they could reason about the conditions under which bulbs light or do not light. The T-chart that served to organize the drawings oriented teachers first to observable features of the circuits (lit, unlit), thus requiring no higher-order inferences about the status of a circuit as complete or incomplete. Had the organizing categories been *complete* and *incomplete*, the learners would have had to move immediately to inferences make the following inference: a circuit is complete because it produced heat, light, or both heat *and* light. Or, they would have had to look at all the connections, and based on their understanding of a complete path or a complete set of contact points, diagnose the circuit as being complete or incomplete. By de-emphasizing the inferences at this stage, the facilitator in Course 1 shifted the burden of data analysis to the whole group rather than placing it on individual contributors. In addition, she created an opportunity assess and respond to the teachers' understanding through the public and collective process of commenting on the observations and constructing inferences and claims. Working with the diagrams displayed on the T-chart, and responding to the ideas expressed by the group, she invited teachers to annotate the observational drawings in ways that highlighted patterns in the observational data (for example, common patterns in the connections that created a complete circuit). As teachers began to generate inferences and claims, the "what we're learning" chart became the repository for them (for example, "you have to have two connection points on the bulb for it to light: at the jacket and at the tip" and "you have to have 2 connection points on the battery for the bulb to light: on the positive side and on the negative side.")

These generalized claims were statements that would apply to all cases of 1 bulb, 1 wire, and 1 battery setups, assuming that all of the materials were working correctly, and would be expected to have predictive power. That is, on the basis of these generalized claims, the teacher/learners would be able to look at various circuit configurations and predict whether the bulb would light.

In Course 1, nearly all of the whole group discussion revolved around the interplay of the evidence provided on the T-chart and the evolving set of claims inventoried on the “what we’re learning” chart — and thus on the central relationship between evidence and inference in the forming of scientific generalization. The pedagogical uses of the inscriptions (supporting conceptual understanding) were thus tightly joined to the scientific practices of reasoning from evidence.

In Courses 2 and 3, the pedagogical choices (what inscriptions were introduced, at what points, and in relationship to what tasks or developments) were less clearly joined to a conception of scientific reasoning. In both cases, however, the diagrams created on the basis of the teachers’ observations were treated as more or less ancillary, essentially points of departure for conceptual discussions to be aided by and anchored to other forms of representation (the Circuit Connections chart and the bulb architecture drawing). And in both cases, the “learning chart” appeared late in the process, primarily to sum up the prior discussion rather than as a device for working back and forth between evidence and inference as a group works toward generalized statements.

By virtue of the T-chart organizer, the participants in Course 2 were better positioned than those in Course 3 to engage in the kinds of scientific reasoning processes evident in Course 1. In Course 2, the T-chart displayed eight circuit diagrams (four lit, four unlit), potentially offering plentiful opportunity for the identification and comparison of distinguishing features and for the development of generalized rules. However, the T-chart was used only to elicit quick summaries of each group’s four diagrams, the first taking a mere 80 seconds, the second —as a result of moves by the facilitator—a little less than 3 minutes. The T-chart was abandoned after only 4 minutes, and subsequent discussion —which remained focused on the conditions for completing a circuit and lighting a bulb — took place in the context of a different representation: the Circuit Connections Chart. Working with that new representation allowed for the comparison of complete, incomplete and short circuits, but did not engage people in moving back and forth from observation to inference in the formation of general understanding. Only when the teachers had completed work on the Circuit Connections chart did the facilitator invite summary statements on a new chart with two organizing headings: connections needed to light a bulb; and connections needed to make a complete circuit. The facilitator concluded the whole group segment by asking teachers to produce a diagrammatic representation or concept map showing their understanding of the main concepts; that activity provided a means for teachers to demonstrate that they understood the session’s fundamental concepts, but the teachers had not had the same experience as those in Course 1 in moving consistently between evidence they had generated and inferences they could support by reference to that evidence.

In Course 3, the absence of the T-chart to organize the teachers’ circuit diagrams resulted in a “data set” on which much organizational work remained to be done before systematic comparisons could readily be made; when that organizational work had been completed after 18 minutes (through serial presentations by each of the ten teachers), the facilitator shifted attention to the Circuit Connections chart. At that point, a co-facilitator proposed to begin recording summary statements on a chart posted on another wall, but those statements (e.g., “all circuits that light are complete circuits,” “not all complete circuits can produce light,” “short circuits are complete circuits”) remained at a high level of abstraction and were never evaluated in light of or explicitly linked to the evidence initially produced by the teachers.

What is difficult to convey in these short summaries, absent the supporting transcript detail, video, or photographs of the successively annotated diagrams and charts, is the degree to which the expert's facilitation in Course 1 session was "layered" to provide multiple cognitive scaffolds, to build norms of collaborative sense-making based in evidence, and to afford substantial scope for the interests and ideas generated by the teachers as learners.

As an example of such layering (further detailed in Wong, 2010), the first 20 minutes of the whole-group discussion in Course 1 focused on elaborating the ideas and relationships represented by only three posted circuit drawings; in that time, the facilitator's moves helped the participants to make the most of the evidence they had displayed as they offered verbal statements and conjectures, annotated the diagrams with writing and other marks to identify salient features, and re-organized the T-chart to highlight the relationships between lit and unlit bulbs and complete and incomplete circuits. Early summary statements or generalizations made by a participant —even if evidently correct and even if offered in response to a facilitator question— were not taken as a sufficient basis for moving on to a next task. In one example, the facilitator responded to an extended teacher-to-teacher exchange by inviting just such a summary: "So, it sounds like, um, there's something about jacket and tip that we're talking about. Can somebody summarize what that is?" When the requested summary statements had been offered by participants and recorded on the learning chart, the facilitator then asked participants to relate the statements back to the available evidence, underscoring the request by moving physically back to the posted diagrams. When participants began to identify evidence contained in the drawings, the facilitator then invited individuals to take a marker and highlight the selected elements of the posted drawing. By the end of the discussion of the drawings, the teachers and facilitator in Course 1 had jointly produced an annotated and re-organized set of drawings that visually highlighted parts, connections and pathways, together with a set of summary statements recorded on the "what we're learning" chart immediately adjacent to the T-chart. By this time, teachers were positioned to offer well-specified as well as accurate predictions and explanations regarding complete and incomplete circuits.

The kind of pedagogical "layering" evident in Course 1 was substantially less evident in the courses facilitated by the newly-recruited facilitators. Yet participants in Courses 2 and 3 did develop a new level of understanding of the key concepts targeted by this course session, as evident both in the teacher knowledge scores and in the session discourse. That is, the analysis suggests a level of cross-course implementation sufficient to produce desired outcomes, but also a level of implementation that might well be enhanced by taking systematic account of the emerging findings in the training and support of facilitators.

## Discussion

Using a randomized experimental design, this study established that three closely related but systematically varied professional development models brought about highly significant gains in teachers' and students' scores on multiple-choice tests of science content knowledge, well beyond those of comparable control groups, and that effects of the courses were maintained a year later. The score increases for students of intervention teachers occurred across a wide range of English language proficiency, from those speaking very little or no English to intermediate English learners, as well as native and non-native fluent-English-proficient students, and student gains did not differ significantly based on sex or race/ethnicity. These findings suggest that all three of the courses have design features that are effective at preparing teachers to support their students' science learning. The three interventions had in common identical collaborative science activities that engaged teachers in investigating and applying elementary grade electric circuits content, and it is not surprising that the three raised teachers' and students' scores on multiple-choice electric circuits tests. This result suggests that important elements in professional development can be configured in a number of ways and still have beneficial effects. The presence of certain characteristics is essential for high quality professional development, but it is possible for professional development to embody them in a variety of effective ways.

However, the three courses differed in key design elements, particularly with respect to sources of pedagogical content knowledge, and their efficacy for producing key outcomes varied accordingly. On a second measure of teacher and student content knowledge, the quality of explanations and applications in response to open-ended test items, the three courses significantly raised teacher explanation scores in Year 1, but only Teaching Cases had significant treatment effects in the follow-up year. For students, in Year 1 only the Looking at Student Work course significantly raised science explanation scores. In that year, Looking at Student Work teachers taught the unit on electric circuits concurrently with taking the professional development course. This meant that the Year 1 students in their electric circuits lessons were completing explanation tasks that were assigned as part of their teachers' Looking at Student Work course shortly before the students took the content posttest, giving them a considerable advantage over students of the other two intervention groups. Interestingly, in the follow-up year, the long-term effects of the Teaching Cases course matched the effects of the Looking at Student Work course, with both significantly raising students' science explanation scores. In contrast, other than teachers' Year 1 results, the Content Immersion course did not raise teachers' or students' scores on science explanations and applications. This pattern likely reflects the fact that Teaching Cases and Looking at Student Work both included analysis of student work and attention to the importance of using classroom tasks that are open-ended enough to elicit useful information about students' conceptual understandings, whereas the Content Immersion course did not include these components.

Teaching Cases and Looking at Student Work also significantly raised teacher pedagogical content knowledge scores, whereas the Content Immersion course did not. The measure of pedagogical content knowledge used in this study scored most highly student actions (e.g, "I would have them build the circuit"), incorporation of instructional representations and activities to help students make sense of phenomena (e.g., "We would trace the flow of current through the wires and into the bulb"), and explication of the specific science learning goals that would be targeted. Teaching Cases and Looking at Student Work both included explicit attention to the science understandings embodied in student work, and in analyzing classroom instruction in relation to helping students improve their understandings. Therefore, the significant impact on pedagogical content knowledge only resulting from Teaching Cases and Looking at Student Work courses is to be expected.

Teachers' ratings of the courses clearly express their preference for the Teaching Cases and Looking at Student Work courses over Content Immersion, in that the focus on content to the exclusion of classroom and student artifacts and reflection limited their opportunities to improve their own practices.

Finally, we also showed that the large impact of the courses on teacher content knowledge only partially accounts for student outcomes. Teachers get something else out of the courses, additional pedagogical knowledge that changes their teaching practices. Our measure of teacher pedagogical content knowledge was found to be a significant predictor both of student multiple-choice test scores and student science explanation scores. When both teacher content knowledge and pedagogical content knowledge were included in the HLM models, content knowledge was a significant predictor of student performance on the multiple-choice test items, but pedagogical content knowledge was not. However, when student science explanation scores were modeled instead of multiple-choice scores, both teacher content knowledge and pedagogical content knowledge were significant predictors of student performance.

### *Implications of the Findings*

The study demonstrates that well-designed professional development of moderate duration—in this case, one five-day course—can make a surprisingly big difference for teaching and learning. We found that it is possible to achieve effect sizes of 2.0 for teachers and 0.7 for students with high quality professional development experiences and a rigorous research design. Few studies to date demonstrate links between teacher professional development and student achievement, especially in science. This randomized controlled study clearly makes this connection, showing positive impacts for both teachers and students, as well as establishing a direct relationship between the two. Furthermore, results indicate significant effects for English learners, despite the considerably smaller size of the English learner sample. The intervention targeted regular science teachers and classes, not pull-out classes taught by specialists in teaching English learners. This suggests that the course has the potential to benefit the large majority of English learners in a way that enhances all students' opportunities to learn.

The current findings provide districts and schools with three research-based options of science professional development that have a proven track record of strengthening teachers' knowledge and practice in ways that benefit their students. The results for the Teaching Cases course are consistent with a previous body of quasi-experimental research on the Understanding Science case-based teacher professional development courses, which had indicated strong treatment effects for elementary grade teachers and students, as well as with a randomized experimental study showing treatment effects at the middle school level. The studies at both grade levels support the conclusion that the model of professional development underlying the courses is strong. For educators, districts, and schools, this is important as it suggests that the entire series of Understanding Science courses may well benefit teachers and students across grade levels and science content areas. If so, the program offers an important step in answering the call from the 2007 report, "Taking Science to School," for coherent, effective science professional development that can be successfully implemented by districts and schools for the benefit of their students.

This study also provides support for the two other professional development models tested, both of which are supported quite often by policy makers and district administrators. Both the Looking at Student Work and the Content Immersion courses led to strong improvements for both teachers and students on multiple-choice tests of science content, improvements that were maintained in the following school year. However, The Content Immersion course did not produce important outcomes that the other courses generated, namely, improvement of teachers' and students' scores on open-ended items measuring explanations and applications of science content, and strengthening of teachers' pedagogical content knowledge as measured in this study. It was also evaluated lowest by teachers both in value and impact.

The fact that Teaching Cases and Looking at Student Work were both so efficacious at raising teacher and student scores and teacher pedagogical content knowledge suggests that they are effective for different reasons, and that professional development that incorporates the essential features of both is likely to produce multiple improvements in teaching and learning. Future investigations of professional development should evaluate these hybrids.

Prior to this study, the authors of this report could find no other published courses that combine science content-specific investigations in conjunction with analysis of student work and pedagogical practice. These distinguishing features of the Understanding Science model warrant further study, to understand the contributions of the various features of the professional development, along with the interplay among the components. In future research, course impact on student achievement should be measured in ways that are both proximal to the intervention (like the topic-specific project-administered tests here) and distal (standardized tests). Additional validation and extension is warranted for the promising measure of pedagogical content knowledge that was developed for use in this study. This measure was sensitive to differences among the professional development courses, and also was a significant predictor of student test scores.

The results provide a step toward supporting the Understanding Science logic model. The model posits that student achievement effects are mediated by improvements in teachers' content knowledge pedagogical content knowledge, and pedagogical practices. To accomplish this, the courses emphasize science content that is situated in activities and scenarios based in student curricula and instruction. The positive effects suggest that policy makers should invest in professional development that builds teachers' in-depth knowledge of the science in the context of classrooms and students, rather than focusing on general pedagogy or purely on content.

### *Strengths and Limitations of the Analyses*

In terms of research design, the power in this study lies in the combination of several design elements. The study compared carefully designed professional development variants, with components in common and that differ, each of which represents a strong option of interest to policy makers. The study used a set of measures driven by a conceptual logic model of the professional development model's target outcomes, and implemented a rigorous randomized experimental design that permits inferences about causal relationships. Finally, rich qualitative data were collected in addition to the test data to illuminate processes and relationships underlying the quantitative patterns.

As described in this report, 175 of the 446 teachers who were randomly assigned to intervention and control groups left the study before data collection was completed, raising concerns about attrition bias. To the extent that these teachers differed from participating teachers, such attrition could reduce external validity (the degree to which the results can be generalized from the retained teacher sample). Such attrition could also bias impact estimates if the attrition is associated with the study outcome measures and if attrition rates differ between intervention and control groups. Based on the analyses of equivalence between the intervention and control groups, there is little evidence of selective attrition.

Since this study recruited a purposively targeted sample, these findings should only be generalized to teachers for whom the tested professional development is a priority. This holds for the original 446 teachers who were recruited and randomly assigned to intervention and control groups, and for the 271 teachers remaining after attrition who provided teacher and/or student data. Since teacher and student participation in the study was voluntary, we cannot quantify whether students unwilling to participate in the science tests would have performed differently from the study samples described in this report.

Finally, the meaning of this study depends upon the validity of the measures used, and two of the measures in this study—the measures of science explanation, and of teacher pedagogical content knowledge—were created for this study without a foundation of measurement development and testing efforts. These measures were good enough to detect differences between interventions, so are promising but need additional validation.

### *The Feasibility of Successful Dissemination*

In this and previous experimental and quasi-experimental studies, the professional development courses were delivered through cadres of staff developers who were trained in face-to-face facilitation academies to lead teacher courses in their regions. The successful outcomes indicate that the courses can be used across multiple sites, by multiple users, suggesting the potential for broad dissemination. The successful train-the-trainer model has the potential for high impact at a relatively low cost and should be scaled up and tested. Analysis of newly-trained facilitators' implementation of the courses has yielded a set of potentially significant improvements to the training model. Since the Understanding Science implementation model focuses on training and supporting facilitation teams, rather than providing direct service to teachers, the effectiveness of facilitator training and support is key to successful replication. The model relies on facilitation academies and written facilitator guides as the key mechanisms of support, and could be enhanced with an online portal for networking, sharing questions and strategies, and attending webinars.

The cost effectiveness of the Understanding Science model of implementation argues further for tests of scaled-up efforts. The model allows for a lean organizational infrastructure that can have a broad reach from the leadership level to classroom teachers and students. Facilitation academies require small numbers of professional development experts (WestEd staff and highly skilled consultants) to lead the trainings with an estimated ratio of 1:16 for each week an academy is offered. These 16 trained facilitators can then go back to their sites and work in pairs to lead teacher courses with an estimated ratio of 2:20 facilitators to teachers. These 20 teachers, with their enhanced knowledge and skill, can go back to their schools and impact an average of 30 students each year they remain in the classroom. With this multiplier effect, a single facilitation academy led by one highly skilled person can indirectly affect the science learning of almost 5,000 students in a single year. Following this logic, it is reasonable to assume the single highly skilled trainer could lead four or more facilitation academies in a year, such that nearly 20,000 students would be affected (or 100,000 at the middle school level where teachers come in contact with 150 students a year). With no additional cost, the same numbers would benefit in subsequent years.

In summary the results suggest that Understanding Science professional development courses effectively address real educational needs, and can be implemented with good fidelity in the existing context of many school systems (e.g., summer institutes) with a relatively small investment of teachers’ time and district resources.

# References

Aaronson, D., Barrow, L., & Sanders, W. (2003). *Teachers and student achievement in the Chicago public high schools.* Chicago: Federal Reserve Bank of Chicago.

American Association for the Advancement of Science. (1993). *Benchmarks for Science Literacy*. New York: Oxford University Press. Available online at: http://www.project2061.org//tools/benchol/bolframe.html

Banilower, E., Boyd, S. E., Pasley, J. D., & Weiss, I. R. (2006). *Lessons from a decade of mathematics and science reform: A capstone report for the local systemic change through teacher enhancement initiative*. Chapel Hill, NC: Horizon Research, Inc.

Barnett, J., and Hodson, D. (2001). Pedagogical context knowledge: Toward a fuller understanding of what good science teachers know. *Science Education, 85*, 426-453.

Barnett-Clarke, C. & Ramirez, A. (in press). Case discussions. In L. B. Easton (Ed.), *Powerful designs for professional development*. Oxford, OH: National Staff Development Council.

Baxter, J. A., & Lederman, N. G. (1999). Assessment and measurement of pedagogical content knowledge. In J. Gess-Newsome & N. G. Lederman (Eds.), *Examining pedagogical content knowledge* (pp. 147-161). Norwell, MA: Kluwer Academic Publishers.

Birman, B., Desimone, L., Porter, A., & Garet, M. (2000). Designing professional development that works. *Educational Leadership, 57*(8), 28-33.

Blank, R.K., de las Alas, N., & Smith, C. (2007). *Analysis of the quality of professional development programs for mathematics and science teachers: Findings from a cross-state study.* Washington DC: CCSSO.

Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher*, 33(8), 3 - 15.

Borko, H. (2006). *Learning from the problem-solving cycle*. Paper prepared for the Symposium on Professional Learning Experiences and the Practice of Teaching: the Role of Artifacts of Practice at the annual meeting of the American Educational Research Association. San Francisco, April 10.

Borko, H., Koellner, K., Jacobs, J., Roberts, S. A., Baldinger, E., & Risley, R. (2010). *Preparing instructional leaders to facilitate mathematics professional development*. Paper presented at the American Educational Research Association.

Boruch, R. F., DeMoya, D., & Snyder, B. (2002). The importance of randomized field trials in education and related areas. In F. Mosteller and R. Boruch (Eds.), *Evidence matters: Randomized trials in education research* (pp. 50-79). Washington, DC: Brookings Institution Press.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How people learn. Brain, mind, experience, and school: Expanded edition*. Washington, DC: National Academy Press.

Brewer, D., & Goldhaber, D. D. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis, 22*, 129-145.

Broughman, S. P., & Rollefson, M. R. (2000). Teacher supply in the United States: Sources of newly hired teachers in public and private schools: 1987-1988 to 1993-94. *Education Statistics Quarterly, 2*(3), 28-32.

Brown, A. L., Bransford, J. D., & Nasir, N. (2000). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.

California Department of Education. (2003). *Science content standards for California public schools: Kindergarten through grade twelve*. Sacramento, CA: Author.

Clermont, C.P., Krajcik, J.S., & Borko, H. (1993). The influence of an intensive in-service workshop on pedagogical content knowledge growth among novice chemical demonstrators. *Journal of Research in Science Teaching, 30*(1), 21-43.

Cobb, P. (2002). Reasoning with tools and inscriptions. *Journal of the Learning Sciences, 11*(2&3), 187-215.

Cohen, D. K., and Hill, H. C. (2000). Instructional policy and classroom performance: The mathematics reform in California. *Teachers College Record, 102*(2): 294–343.

Cohen, D. K., & Hill, H. C. (2001). *Learning policy: When state education reform works*. New Haven: Yale University Press.

Cook, T. (2001). Sciencephobia. *Education Next 1*(3), 62-68.

Cook, T. D., Habib, F.-N., Phillips, M., Settersten, R. A., Shagle, S. C., and Degirmencioglu, S. M. (1999). Comer's School Development Program in Prince George's County, Maryland: A theory-based evaluation. *American Educational Research Journal, 36*(3), 543-597.

Daehler, K. R., & Shinohara, M. (2001). A complete circuit is a complete circle: Exploring the potential of case materials and methods to develop teachers' content knowledge and pedagogical content knowledge of science. *Research in Science Education, 31*(2), 267-288.

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher, 38*, 181-199.

Desimone, L. M., Porter, A. C., Garet, M. S., Yoon, K. S., & Birman, B. F. (2002). Effects of professional development on teachers' instruction: Results from a three-year longitudinal study. Educational Evaluation and Policy Analysis, 24, 81–112.

Duschl, R. A., Schweingruber, H.A., and Shouse, A.W. (Eds.) (2007). *Taking Science to School: Learning and Teaching Science in Grades K-8.* Washington, D.C.: National Academies Press.

Elliott, R., Kazemi, E., Lesseig, K., Mumme, J., Carroll, C., Kelley-Petersen, M., et al. (2009). Conceptualizing the work of leading mathematical tasks in professional development. *Journal of Teacher Education, 60*(4), 364-379.

Erickson, F. (1992). Ethnographic microanalysis of interaction. In M. D. LeCompte, W. L. Millroy, & J. Preissle (Eds.), *The handbook of qualitative research in education* (pp. 202-225). San Diego: Academic Press, Inc.

Falk, A. (2009b). *Teachers learning from professional development in elementary science: Reciprocal relationships between formative assessment and pedagogical content* *knowledge.* Unpublished Ph.D. dissertation, University of California, Berkeley.

Fennema, E., Carpenter, T. P., Franke, M. L., Levi, L., Jacobs, V. R., & Empson, S. B. (1996). A longitudinal study of learning to use children's thinking in mathematics instruction. *Journal for Research in Mathematics Education, 27*, 403-434.

Fishman, J. J., Marx, R. W., Best, S., & Tal, R. T. (2003). Linking teacher and student learning to improve professional development in systemic reform. *Teaching and Teacher Education, 19*, 643–658.

Foorman, B. R., Francis, D. J., Fletcher, J. M., Schatschneider, C., & Mehta, P. (1998). The role of instruction in learning to read: Preventing reading failure in at-risk children. *Journal of Educational Psychology, 90*(1), 37-55.

Franke, M. L., Carpenter, T., Levi, L., & Fennema, E. (2001). Capturing teachers' generative change: A follow-up study of professional development in mathematics. *American Educational Research Journal, 38*, 653-690.

Fulp, S. L. (2002). *2000 National Survey of Science and Mathematics Education: Status of Elementary School Science Teaching.* Chapel Hill, NC: Horizon Research, Inc. Retrieved from http://2000survey.horizon-research.com/reports/elem_science.php.

Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal, 38*(4), 915-945.

Geddis, A.N., Onslow, B., Beynon, C., & Oesch, J. (1993). Transforming content knowledge: Learning to teach about isotopes. *Science Education, 77*(6), 575-591.

Goodwin, C. (1994). Professional vision. *American Anthropologist, 96*(3), 606-633.

Hall, R. (2000). Video recording as theory. In D. Lesh & A. Kelley (Eds.*) Handbook of research design in mathematics and science education* (pp. 647-664). Mahweh, NJ: Lawrence Erlbaum.

Hanssen, C. E. (2006, October). *The Milwaukee Mathematics Partnership: A path model for evaluating teacher and student effects.* Paper presented at the MSP Evaluation Summit II, Minneapolis, MN.

Hapgood, S., Magnusson, S. J., & Palincsar, A. S. (2004). Teacher, text, and experience: A case of young children's scientific inquiry. *Journal of the Learning Sciences, 13*(4), 455-505.

Hart, J. E.& & Lee, O. (2003). Teacher professional development to improve the science and literacy achievement of English language learners. *Bilingual Research Journal, 27*(3), 475-501.

Heller, J. I., Daehler, K., and Shinohara, M. (2003). Connecting All the Pieces: Using an Evaluation Mosaic to Answer an Impossible Question. *Journal of Staff Development, 24*: 36-41.

Heller, J. I., Daehler, K. R., Shinohara, M., & Kaskowitz, S. R. (2004, April). *Fostering pedagogical content knowledge about electric circuits through case-based professional development*. Paper presented at the annual meeting of the National Association for Research on Science Teaching, Vancouver, B.C., Canada.

Heller, J. I., & Kaskowitz, S. R. (2004). *Final evaluation report for Science Cases for Teacher Learning: Impact on teachers, classrooms, and students, Project years 2000 – 2003.* Technical report submitted to Stuart Foundation.

Heller, J. I., Shinohara, M., Miratrix, L., Rabe-Hesketh, S., and Daehler, K. R. (2010, March). *Learning Science for Teaching: Effects of Professional Development on Elementary Teachers, Classrooms, and Students.* Paper presented at the 2010 Conference of the Society for Research on Educational Effectiveness, Washington, D.C.

Hewson, P. W., Kahle, J. B., Scantlebury, K., & Davis, D. (2001). Eq uitable science education in urban middle schools: Do reform efforts make a difference? *Journal of Research in Science Teaching, 38*(10), 1130–1144.

Hill, H. C. (2007). Learning in the teaching workforce. *The Future of Children, 17*(1), 111–127.

Hill, H. C. Rowan, B., and Ball, D. L. (2005). Effects of Teachers' Mathematical Knowledge for Teaching on Student Achievement. *American Educational Research Journal, 42*(2): 371-406.

Hindmarsh, J., & Heath, C. (1998). Video and the analysis of objects in action. *Communication & Cognition, 31*(2/3), 111-130.

Jacob, R., Zhu, P., & Bloom, H. (2010). New empirical evidence for the design of group randomized trials in education. *Journal of Research on Educational Effectiveness, 3*(2), 157–198.

Kazemi, E., & Franke, M. L. (2004). Teacher learning in mathematics: Using student work to promote collective inquiry. *Journal of Mathematics Teacher Education, 7,* 203-235.

Kesidou, S., & Roseman, J. (2001). How well do middle school science programs measure up? Findings from Project 2061's curriculum review. *Journal of Research in Science Teaching, 36*(6), 522-549.

Knapp, M. S., McCaffrey, T., and Swanson, J. (2003, April). *District Support for Professional Learning: What Research Says and Has Yet to Establish.* Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Latour, B. (1990). Drawing things together. In M. Lynch & S. Woolgar (Eds.), *Representation in scientific practice* (pp. 19-68). Cambridge, MA: MIT press.

Latour, B. & Woolgar, S. (1986). *Laboratory life: The social construction of scientific facts.* Princeton, NJ: Princeton University Press.

Lederman, N.G., & Gess-Newsome, J. (1992) Do subject matter knowledge, pedagogical knowledge and pedagogical content knowledge constitute the ideal gas law of science teaching? *Journal of Science Teacher Education, 3*(1), 16-20.

Lederman, N.G., Gess-Newsome, J., & Latz, M.S. (1994). The nature and development of preservice science teachers' conceptions of subject matter and pedagogy. *Journal of Research in Science Teaching, 31*(2), 129-146.

Lee, O. (2002). Science inquiry for elementary students from diverse backgrounds. In W. G. Secada (Ed.), *Review of Research in Education,* Vol. 26 (pp. 23-69). Washington, DC: American Educational Research Association.

Lee, O., & Fradd, S. H. (2001). Instructional congruence to promote science learning and literacy development for linguistically diverse students. In D. R. Lavoie & W-M. Roth (Eds.), *Models for science teacher preparation: Bridging the gap between research and practice*. (pp. 109-126). Dordrecht, the Netherlands: Kluwer Academic Publishers.

Lemke, J. L. (1998). Multiplying meaning: Visual and verbal semiotics in scientific text. In J. R. Martin & R. Veel (Eds.), *Reading science* (pp. 87-113). London: Routledge.

Little, J. W. (2002). Locating learning in teachers' communities of practice: Opening up problems of analysis in records of everyday work. *Teaching and Teacher Education, 18*(8), 917-946.

Little, J. W. (2003). Inside teacher community: Representations of classroom practice. *Teachers College Record*, *105*(6), 913-945.

Little, J. W. (2006). *NEA Research Best Practices: Professional Community and Professional Development in the Learning-Centered School.* Washington, D.C.: National Education Association. Retrieved October 23, 2010, from: http:// www.nea.org /tools/30380.htm

Little, J.W., Gearhart, M., Curry, M., & Kafka, J. (2003). Looking at student work for teacher learning, teacher community and school reform. *Phi Delta Kappan, 85*(3), 184-192.

Loughran, J., Mulhall, P., & Berry, A. (2003). In search of pedagogical content knowledge in science: Developing ways of articulating and documenting professional practice. *Journal of Research in Science Teaching, 41*(4), 370-291.

Lynch, M. (1990). The externalized retina: Selection and mathematization in the visual documentation of objects in the life sciences. In M. Lynch & S. Woolgar (Eds.), *Representation in scientific practice* (pp. 153-186). Cambridge, MA: MIT Press.

Magnusson, S., Krajcik, J., & Borko, H. (1999). Nature, sources, and development of pedagogical content knowledge for science teaching. In Gess-Newsome, J. & Lederman, N.G. (Eds.) *Examining pedagogical content knowledge* (pp. 95-132). Washington, D.C.: National Academy Press.

Maxwell, J. A. (2004). Causal explanation, qualitative research, and scientific inquiry in education. *Educational Researcher*, 33(2), 3-11.

Merseth, K. K. (1996). Cases and case methods in teacher education. In J. Sikula (Ed.), *Handbook of research on teacher education* (pp. 722-744). New York: MacMillan Publishing Company.

Monk, D. H. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review, 13*, 125-145.

Monk, D. H., & King, J. (1994). Multilevel teacher resource effects on pupil performance in secondary mathematics and science. In R. G. Ehrenberg (Ed.), *Choices and consequence.* Ithaca, NY: ILR Press.

National Research Council (1996). *National Science Education Standards.* Washington, D.C.: National Academies Press. Retrieved October 23, 2010, from http://www.nap.edu/catalog.php?record_id=4962.

National Research Council. (2000). *Educating teachers of science, mathematics and technology: New practices for the new millennium* (Committee on Science and Mathematics Teacher Preparation). Washington, DC: National Academy Press.

National Science Foundation. (2004). National Science, Technology, Engineering, and Mathematics Education Digital Library (NSDL)—Program solicitation. Arlington, VA: Author. Available at http://www.nsf.gov/pubsys/ods/getpub.cfm?ods_key=nsf04542 .

National Staff Development Council (2001). *National Staff Development Council's Standards for Staff Development (Revised)*. Retrieved October 23, 2010, from http://www.learningforward.org/standards/index.cfm.

O'Reilly, F. E., & Weiss, C. H. (2007). *Opening the black box. Using theory-based evaluation to understand professional development for K–12 teachers of math and science.* Paper presented at the American Educational Research Association, April 2006.

Parker, J. & Heywood, D. (2000) Describing the relationship between subject knowledge and pedagogic content knowledge in primary teachers' learning about forces. *International Journal of Science Education, 22*(1), 89-111.

Partnership for 21st Century Skills (2008). *21st Century Skills, Education and Competitiveness: A Resource and Policy Guide.* Tucson, AZ: Author. Retrieved October 23, 2010, from http://www.p21.org/index.

Radinsky, J., Goldman, S., & Singer, M. (2008). Students' sense-making with visual data in small-group argumentation. In *Proceedings of the International Conference of the Learning Sciences* (June 2008).

Redish, E. F. (1996). *Discipline-specific science education and educational research: The case of physics*. Paper prepared for the Committee on Developments in the Science of Learning for the Sciences of Science Learning: An Interdisciplinary Discussion. Department of Physics and Astronomy, University of Maryland.

Rivers, J. C., & Sanders, W. L. (2002). In L. T. Izumi & M. E. Williamson (Eds.), *Teacher quality* (pp. 13-23). Stanford, CA: Hoover Institution Press.

Rodriguez, A. J., (2004). *Turning despondency into hope: Charting new paths to improve students' achievement and participation in science education.* Tallahasse, FL: The Southeast Eisenhower Regional Consortium for Mathematics and Science Education at SERVE.

Rosebery, A., B. Warren and F. Conant (1992). Appropriating scientific discourse: Findings from language minority classrooms. *The Journal of the Learning Sciences, 2,* 61-94.

Rowan, B., Chiang, F. S., & Miller, R. J. (1997). Using research on employee's performance to study the effects of teacher on students' achievement. *Sociology of Education, 70,* 256-284.

Saxe, G. B., Gearhart, M., & Nasir, N. (2001). Enhancing students' understanding of mathematics: A study of three contrasting approaches to professional support. *Journal of Mathematics Teacher Education, 4*(55-79).

Sawchuk, S. (2010, November). Professional development for teachers at crossroads. *Education Week, 30*(11), s2-s4. Retrieved from http://www.edweek.org Date Accessed: 11/10/2010 9:33:01 AM.

Scher, L., & O'Reilly, F. (2009). Professional development for K–12 math and science teachers: What do we really know? *Journal of Research on Educational Effectiveness, 2*(3), 209–249.

Shinohara, M., Daehler, K. R., & Heller, J. I. (2004, April). *Using a pedagogical content framework to determine the content of case-based teacher professional development in science.* Paper presented at the annual meeting of the National Association for Research in Science Teaching, Vancouver, B.C., Canada.

Shulman, J. H. (Ed.). (1992). *Case methods in teacher education*. New York: Teachers College Press.

Shulman, L. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4-14.

Shulman, L.S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Education Review, 57*(1), 1-22.

Shymansky, J. and C. Matthews (1993). Focus on children's ideas about science: An integrated program of instructional planning and teacher enhancement from the constructivist perspective. *The Proceedings of the Third International Seminar on Misconceptions and Educational Strategies in Science and Mathematics*. Ithaca, NY: Misconceptions Trust.

Slavin, R. E. (2002). Evidence-based education policies: Transforming educational practice and research. *Educational Researcher, 31*(7), 15–21.

Smith, D.C. & Neale, D.C. (1989). The construction of subject-matter knowledge in primary science. *Teaching and Teacher Education, 5,* 1-20.

Spillane, J. P., Diamond, J. B., Walker, L. J., Halverson, R., & Jita, L. (2001). Urban school leadership for elementary science instruction: Identifying and activating resources in an undervalued school subject. *Journal of Research in Science Teaching, 38*(8), 918–940.

Stevens, R. & Hall, R. (1998). Disciplined perceptions: learning to see in technoscience. In M. Lampert and M. Blunk (Eds.), *Talking mathematics in school: Studies of teaching and learning* (pp. 107-149). Cambridge, UK: Cambridge University Press.

Terrell, Nicholas. (2007). STEM Occupations: High-Tech Jobs for a High-Tech Economy. *Occupational Outlook Quarterly.* Washington, DC: Office of Occupational Statistics and Employment Projections, U.S. Department of Labor.

Torres, H.N., & Zeidler, D.L. (2002) The effects of English language proficiency and scientific reasoning skills on the acquisition of science content knowledge by Hispanic English language learners and native English language speaking students. *Electronic Journal of Science Education, 6*(3): Article Four.

U.S. Department of Education (2004). *Trends in International Mathematics and Science Study: Science items.* Washington, DC: U.S. Department of Education, National Center for

Educational Evaluation. Retrieved August 13, 2010 from http://nces.ed.gov/timss/educators.asp

Van Driel, J.H., Verloop, N., & De Vos, W. (1998). Developing science teachers' pedagogical content knowledge. *Journal of Research in Science Teaching, 35*(6), 673-695.

Wallace, M. R. (2009). Making Sense of the Links: Professional Development, Teacher Practices, and Student Achievement**.** *Teachers College Record, 111*(2): 573-596. Retrieved from http://www.tcrecord.org ID Number: 15224, Date Accessed: 11/9/2010 3:34:37 PM.

Wasserman, S. (1993). *Getting down to cases: Learning to teach with case studies.* New York: Teachers College Press.

Weiss, I. R., Banilower, E. R., McMahon, K. C., and Smith, P. S. (2001). *Report of the 2000 National Survey of Science and Mathematics Education.* Chapel Hill, NC: Horizon Research, Inc.

Weiss, I. R., & Miller, B. (2006). *Developing strategic leadership for district-wide improvement of mathematics education*. Lakewood, CO: National Council of Supervisors of Mathematics.

Wilson, S.M., & Berne, J. (1999). Teacher learning and the acquisition of professional knowledge: An examination of research on contemporary professional development. In A. Iran-Nejad and P.D. Pearson (Eds.). *Review of Research in Education, 23,* 217-234. Washington D.C.: American Educational Research Association.

What Works Clearinghouse. (2008). *Procedures and Standards Handbook (Version 2.0).* Retrieved October 23, 2010, from http://ies.ed.gov/ncee/wwc/pdf/wwc_procedures_v2_standards_handbook.pdf

Wong, N. (2010). Supporting scientific reasoning and conceptual understanding through the use of inscriptions. Unpublished Ph.D. dissertation. Berkeley: University of California, Berkeley.

Wright, S.P., Horn, S.P., & Sanders, W.L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education, 11,* 57-67.